

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/82, 5/10, A01H 5/00, C12Q 1/68</b>		A1	(11) International Publication Number: <b>WO 99/07866</b> (43) International Publication Date: 18 February 1999 (18.02.99)
<p>(21) International Application Number: <b>PCT/US98/16344</b></p> <p>(22) International Filing Date: 5 August 1998 (05.08.98)</p> <p>(30) Priority Data: 60/066,118 6 August 1997 (06.08.97) US</p> <p>(71) Applicant (<i>for all designated States except US</i>): NORTH CAROLINA STATE UNIVERSITY [US/US]; 1 Holladay Hall, Campus Box 7003, Raleigh, NC 27695-7003 (US).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (<i>for US only</i>): MICHALOWSKI, Susan, Margaret [US/US]; 316 Hallwood Court, Holly Springs, NC 27540 (US). SPIKER, Steven, L. [US/US]; 1322 Canterbury Road, Raleigh, NC 27608 (US).</p> <p>(74) Agents: BENNETT, Virginia, C. et al.; Myers, Bigel, Sibley, &amp; Sajovec, P.A., P.O. Box 37428, Raleigh, NC 27627 (US).</p>		<p>(81) Designated States: AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), EE, EE (Utility model), ES, FI, FI (Utility model), GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b>  <i>With international search report.</i>  <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
<p>(54) Title: MATRIX ATTACHMENT REGIONS</p> <p>(57) Abstract</p> <p>Matrix attachment regions isolated from higher plants, and DNA constructs and vectors containing such matrix attachment regions, are described. A method of identifying matrix attachment regions is provided.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## MATRIX ATTACHMENT REGIONS

### Field of the Invention

The present invention relates to matrix attachment regions isolated from a higher plant, and to methods for isolating matrix attachment sequences.

5

### Background of the Invention

- The proteinaceous nuclear 'matrix' or 'scaffold' in the cell nucleus plays a role in determining chromatin structure. Electron micrographs show that nuclear DNA is attached to this scaffold at intervals to produce a series of loops (Zlatanova and Van Holde, *J. Cell Sci.* 103:889 (1992)). Matrix Attachment Regions (MARs; also referred 10 to as scaffold attachment regions or SARS) are genomic DNA sequences which bind specifically to components of the nuclear matrix. See Boulikas, *J. Cell. Biochem.* 52:14 (1993). These sequences are thought to define independent chromatin domains through their attachment to the nuclear matrix. Both transcription and replication are thought to occur at the nuclear matrix.
- 15 Transformation of a cell using a transgene flanked by one or more MARs has been shown to increase expression of the transgene product, compared to transformation using a construct lacking MARs. See Allen et al., *Plant Cell* 8:899 (1996); Bonifer et al., *EMBO J.* 9:2843 (1990); McKnight et al., *Proc. Natl. Acad. Sci. USA* 89:6943 (1992); Phi-Van et al., *Mol. Cell. Biol.* 10:2303 (1990)). Flanking a GUS 20 reporter gene with yeast MARs has been reported to result in higher and less variable transgene expression in plant cells. Allen et al. *Plant Cell* 5:603 (1993).

### Summary of the Invention

In view of the foregoing, a first aspect of the present invention is an isolated DNA molecule having a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, 2, 4-11 and 13, sequences that hybridize to this isolated DNA under 5 stringent conditions.

A further aspect of the present invention is a DNA construct comprising a transcription initiation region, a structural gene operatively associated with the transcription initiation region, and at least one matrix attachment region of the present invention positioned either 5' to the transcription initiation region or 3' to the 10 structural gene.

A further aspect of the present invention is a vector comprising a DNA construct as described above, including plasmids, viruses and plant transformation.

A further aspect of the present invention is a host cell containing a DNA construct as described above, including plant and animal host cells.

15 A further aspect of the present invention is a method of identifying matrix attachment regions in a DNA molecule of known nucleotide sequence, by identifying a sequence section of at least twenty contiguous nucleotides that is at least 90% A or T nucleotides. The method may further comprise preparing a MAR molecule of at least about 300 nucleotides, comprising the identified MAR motif.

20

### Brief Description of the Drawings

Figure 1 provides maps depicting generalized plasmids from the cloning of random matrix associated DNA into pBluescript II SK+. 1A is DNA isolated after digestion with RsaI and ligated into the EcoRV site (clones 1-4, 6-8, 2-23 and 26-28). 1B is DNA isolated after digestion with TaqI and ligated into the Clal site (clones 11, 15, 34 and 35). 1C is DNA isolated after digestion with EcoRI and ligated into the EcoRI site (clones 109, 113, 115 and 116). 1D is DNA isolated after digestion with HindIII and ligated into the HindIII site (clones 201-203, 205, 206, 209, 211 and 216-220). 1E is DNA isolated after digestion with DnaseI and ligated into the HincII site (clones 302, 203, 205, 311 and 319). 1F is a map of ToRB7-6 which serves as a 30 positive control in the exogenous binding assay.

**Figure 2** provides plasmid maps of specific clones or subclones chosen for sequencing. Inserts are indicated by shaded boxes labeled either 'MAR' for binding clones or 'insert' for non-binding clones. **2A** is plasmid pS1 containing MAR SEQ ID NO:1; **2B** is plasmid pS4 containing MAR SEQ ID NO:2; **2C** is plasmid pS8 containing 5 non-binding SEQ ID NO:3; **2D** is plasmid pS115 containing MAR SEQ ID NO:4; **2E** is MAR plasmid pS116; **2F** is plasmid pS116-1.1 containing MAR SEQ ID NO:5, which is a smaller core binding fragment of clone 116; **2G** is MAR plasmid pS202; **2H** is plasmid pS202-1 containing MAR SEQ ID NO:6, which is one of two binding fragments 10 of clone 202; **2I** is plasmid pS202-2 containing MAR SEQ ID NO:7, which is the second of two binding fragments of clone 202; **2J** is plasmid pS203 containing a non-binding insert; **2K** is MAR plasmid pS205; **2L** is plasmid pS205-2 containing MAR SEQ ID NO:8, which is a core binding sequence from clone 205; **2M** is MAR plasmid pS206; **2N** is plasmid pS206-1 containing MAR SEQ ID NO:9, which is a core binding 15 sequence from clone 206; **2O** is MAR plasmid pS211; **2P** is plasmid pS211-1 containing MAR SEQ ID NO:10, which is a core binding sequence from clone 211; **2Q** is MAR plasmid pS217; **2R** is plasmid pS217-1 containing MAR SEQ ID NO:11, which is a core binding sequence from clone 217; **2S** is plasmid pS218 containing non-binding insert SEQ ID NO:12; **2T** is MAR plasmid pS220; **2U** is plasmid pS220-1 containing MAR SEQ ID NO:13, which is a core binding sequence from clone 220; **2V** 20 is plasmid pRB7-6, containing the MAR ToRB7-6 fragment (SEQ ID NO:20) used as a positive control; **2W** is plasmid pGCA887 (containing an insert from yeast ARS1 cloned into the vector pBCKS+ (Stratogene)), which serves as a standard for weak binding to the nuclear matrix (SEQ ID NO:21).

**Figure 3** provides the sequences of the MAR clones and subclones of the 25 present invention, and the control sequences of the known TobRB7 MAR (SEQ ID NO:20) and yeast ARS1 MAR (SEQ ID NO:21).

**Figure 4** provides graphic representations depicting the locations of different MAR DNA motifs within the sequenced clones or subclones. Binding strengths are indicated on a scale of 0-100. The A box (AATAAA/YAAA) (SEQ ID NO:14) is 30 represented by "A", with 8/10 matches required for motif identification. The T box (TTWTWT/TWTT) (SEQ ID NO:15) is represented by "T", with 9/10 matches required for motif identification. The ARS consensus sequence (WTTTATRTTTW) (SEQ ID

NO:16) is represented by 'R', with 10/11 matches required for motif identification. The topoisomerase II consensus sequence (GTNWAYATTNATNNR) (SEQ ID NO:17) is represented by 'O', with 13/15 matches required for motif identification. If motifs overlapped, only one is shown. Filled boxes indicate stretches of 20 base pairs consisting of  $\geq$  90% AT DNA. Base unwinding regions are represented by 'U' (AATATATTT; SEQ ID NO:22; Bode et al., *Science* 255:195 (1992)).

Figure 5 graphs the numbers of blocks of 20 or more nucleotides that consist of 90% or greater A or T nucleotides found in the sequenced clones (SEQ ID NOS: 1, 2, 4-11 and 13) versus binding strength (indicated as between 0-100). The two well-characterized MARs (TobRB7 and ARS1), as well as two non-binding clones (SEQ ID NOs:3 and 12) were included in the analysis.

Figure 6 graphs the %AT found in the sequenced clones (SEQ ID NOS: 1, 2, 4-11 and 13) versus binding strength (indicated as between 0-100). The two well-characterized MARs (TobRB7 and ARS1), as well as two non-binding clones (SEQ ID NOs:3 and 12) were included in the analysis.

Figure 7 graphs the number of T boxes found in the sequenced clones (SEQ ID NOS: 1, 2, 4-11 and 13) versus binding strength (indicated as between 0-100). The two well-characterized MARs (TobRB7 and ARS1), as well as two non-binding clones (SEQ ID NOs:3 and 12) were included in the analysis.

Figure 8 graphs the number of A boxes found in the sequenced clones (SEQ ID NOS: 1, 2, 4-11 and 13) versus binding strength (indicated as between 0-100). The two well-characterized MARs (TobRB7 and ARS1), as well as two non-binding clones (SEQ ID NOs:3 and 12) were included in the analysis.

Figure 9 graphs the number of base unwinding regions (BUR; SEQ ID NO:22) found in the sequence clones (SEQ ID NOS: 1, 2, 4-11 and 13) versus binding strength (indicated as between 0-100). The two well-characterized MARs (TobRB7 and ARS1), as well as two non-binding clones (SEQ ID NOs:3 and 12) were included in the analysis.

Figure 10 graphs the length of the sequenced clones (SEQ ID NOS: 1, 2, 4-11 and 13) versus binding strength (indicated as between 0-100). The two well-characterized MARs (TobRB7 and ARS1), as well as two non-binding clones (SEQ ID NOs:3 and 12) were included in the analysis.

Figure 11 graphs the number of ARS motifs found in the sequenced clones (SEQ ID NOS: 1, 2, 4-11 and 13) versus binding strength (indicated as between 0-100). The two well-characterized MARs (TobRB7 and ARS1), as well as two non-binding clones (SEQ ID NOs:3 and 12) were included in the analysis.

5       Figure 12 graphs the number of Topoisomerase motifs found in the sequenced clones (SEQ ID NOS: 1, 2, 4-11 and 13) versus binding strength (indicated as between 0-100). The two well-characterized MARs (TobRB7 and ARS1), as well as two non-binding clones (SEQ ID NOs:3 and 12) were included in the analysis.

10

#### Detailed Description of the Invention

Matrix attachment regions (MARs) are structural components of chromatin that form topologically constrained loops of DNA through their interaction with the proteinaceous nuclear matrix. MARs have been found to co-localize with a variety of functional elements within the nucleus including transcriptional domain boundaries (Jarman and Higgs, *EMBO J.* 7:3337 (1988); Phi Van and Sträling, *EMBO J.* 7:655 (1988); Levy-Wilson and Fortier, *J. Biol. Chem.* 264:21196 (1990)), promoters, enhancers (Gasser and Laemmli, *Cell* 46:521 (1986); Cockerill and Garrard, *Cell* 44:273 (1986); van der Geest, *Plant J.* 6:413 (1994)), introns (Kas and Chasin, *J. Mol. Biol.* 194:677 (1987); Forrester et al., *Science* 265:1221 (1994)) and putative origins of replication (Brylawski et al., *Cancer Res.* 53:3865 (1993)), suggesting that MARs may play functional roles in addition to their purely structural role within the nucleus. It appears that not all MARs are involved in the same processes, and that categories or groups of these elements with distinct features and functions exist.

25       The characteristics of MARs that dictate their binding to the nuclear matrix are not known. Presently, the only definition of a MAR is operational, based on the ability to bind to the nuclear matrix. It is known that MARs are AT rich, but not all AT-rich DNA will bind to the nuclear matrix. MARs have also been reported to contain a number of short sequence motifs, but the necessity of these motifs has not been established. Motifs reported to occur in MARs include A boxes, T boxes, the ARS consensus and the consensus sequence for *Drosophila* topoisomerase. In addition,

several secondary structure motifs have been reported to be associated with MARs including base pair unwinding regions, bent DNA and single stranded regions.

To date most matrix attachment regions have been identified through their association with a well-characterized gene. This type of sampling creates a bias  
5 that could hinder efforts in defining MAR sequences. It would be useful to be able to identify MARs by sequence alone. The present inventors obtained a group of DNA fragments that were MARs by operational definition, by purifying DNA associated with tobacco NT-1 nuclear matrices prepared using several different nucleases. These sequences were cloned and tested for their ability to rebind to the nuclear matrix, in  
10 order to identify MARs. Once MARs were identified, they were sequenced and analyzed for AT content and the presence of common motifs. The significance of each identified motif was assessed through correlation with the binding strength of MARs to the nuclear matrix.

The present inventors identified a number of novel MAR sequences, and  
15 identified a new MAR motif whose frequency significantly correlates with the binding strength of a MAR. The present inventors found no significant correlation between binding strength and the length of the MAR fragment. However, a significant relationship between binding strength and overall AT content was identified. This is the first report of a correlation between the abundance of certain MAR related motifs and  
20 MAR binding strength. In addition, the newly identified MAR related motif of local AT rich regions (sections of 20 contiguous nucleotides that are  $\geq 90\%$  A and/or T), has a higher correlation to MAR binding strength than any of the previously identified motifs. These findings provide a method for the identification of MAR regions in DNA molecules of known nucleotide sequence. The method comprises identifying, in the  
25 known DNA sequence, regions or areas of the sequence which are at least 20 contiguous nucleotides in length and which consist of at least 90% A and/or T nucleotides. The presence of a 20-bp region of  $\geq 90\%$  AT indicates a MAR; a MAR may contain multiple regions of  $\geq 90\%$  AT. The identification of such regions may be carried out by techniques that are well-known in the art, including sequencing the DNA to be screened  
30 and reviewing a printed DNA sequence for such regions. Contiguous fragments of the original DNA sequence that are from one to several kilobases (from about 3,000

nucleotides, 2,000 nucleotides, or about 1,000 nucleotides) in length to about 500, 400, or 300 bases in length, and which encompass the 20-bp regions of  $\geq$  90% AT can then be isolated (or created *de novo* by known synthesis techniques) and utilized as MARs. Optionally, the isolated fragments can first be tested for MAR binding strength, for example using an exogenous nuclear matrix binding assay as described herein.

The identification of such regions may be carried out by techniques that are well-known in the art, including sequencing the DNA to be screened and reviewing the printed DNA sequence for such regions. Fragments of the original DNA sequence that are from several kilobases in length to about 500, 400, or 300 bases in length, and which encompass the 20-bp regions of  $\geq$  90% AT can then be isolated (or created *de novo* by known synthesis techniques) and utilized as MARs. Optionally, the isolated fragments can first be tested for MAR binding strength, for example using an exogenous nuclear matrix binding assay as described herein.

MARs in nature are double-stranded genomic DNA molecules. The MARs of the present invention include those of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:13. The sequences provided represent one strand of the double-stranded MAR DNA; the sequence of the complementary strand is readily apparent to those of ordinary skill in the art.

It will be apparent to those of skill in the art that minor sequence variations from the sequences provided above will not affect the function of the MARs of the present invention. MAR DNA sequences of the present invention include sequences that are functional MARs which hybridize to DNA sequences of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 or SEQ ID NO:13 (or the complementary sequences thereto) under stringent conditions. For example, hybridization of such sequences may be carried out under conditions represented by a wash stringency of 0.3M NaCl, 0.03M sodium citrate, and 0.1% SDS at 60°C, or even 70°C, in a standard *in situ* hybridization assay. (See J. Sambrook et al., Molecular Cloning, A Laboratory Manual (2d ed. 1989)(Cold Spring Harbor Laboratory)). In general, DNA sequences that act as MARs and hybridize to the DNA sequences given above will have at least 70%, 75%, 80%, 85%, 90%, 95% or even 97% or greater sequence similarity to the MAR sequences.

provided herein. (Determinations of sequence similarity are made with the two sequences aligned for maximum matching; gaps in either of the two sequences being matched are allowed in maximizing matching.)

MARs of the present invention may consist of or comprise the specific sequences provided herein, or nucleotide sequences having substantial sequence similarity to the sequences provided herein that retain MAR functions. As used herein, 'substantial sequence similarity' means that DNA which have slight and non-consequential sequence variations from the specific sequences disclosed herein are considered to be equivalent to the disclosed sequences. In this regard, 'slight and non-consequential' sequence variations mean that sequences with substantial sequence similarity will be functionally equivalent to the sequences disclosed and claimed herein. Functionally equivalent sequences will function in substantially the same manner as the sequences disclosed and claimed herein.

DNA constructs of the present invention may be used to transform cells from a variety of organisms, including animal and plants (*i.e.*, vascular plants). As used herein, plants includes both gymnosperms and angiosperms (*i.e.*, monocots and dicots). As used herein, animals includes mammals, both primate and non-primate. Transformation according to the present invention may be used to increase expression levels of transgenes in stably transformed cells. Cells may be transformed while in cell culture; while *in vivo* or *in situ* in a tissue, organ, or intact organism.

The term "operatively associated," as used herein, refers to DNA sequences on a single DNA molecule which are associated so that the function of one is affected by the other. Thus, a transcription initiation region is operatively associated with a structural gene when it is capable of affecting the expression of that structural gene (*i.e.*, the structural gene is under the transcriptional control of the transcription initiation region). The transcription initiation region is said to be "upstream" from the structural gene, which is in turn said to be "downstream" from the transcription initiation region.

DNA constructs, or "expression cassettes," of the present invention preferably include, 5' to 3' in the direction of transcription, a first matrix attachment region, a transcription initiation region, a structural gene operatively associated with the transcription initiation region, a termination sequence including a stop signal for RNA

polymerase and a polyadenylation signal for polyadenylation (e.g., the nos terminator), and a second matrix attachment region. All of these regions should be capable of operating in the cells to be transformed. The termination region may be derived from the same gene as the transcription initiation or promoter region, or may be derived from  
5 a different gene.

The matrix attachment regions (or "MARs") of the present invention have a nucleotide sequence selected from the group consisting of SEQ ID NOS: 1, 2 4-11 and 13 provided herein. These MARs may be isolated from natural sources or may be chemically synthesized.

10 MARs are known to act in an orientation-independent manner. Poljak et al., *Nucleic Acids Res.* 22:4386 (1994). Genetic constructs of the present invention may contain MARs oriented in either direction (5'-3' or 3'-5'), as direct repeats in a single orientation (→→), direct repeats in the opposite orientation (←←), or either of two possible indirect repeats (←→) or (→←). The genetic constructs of the present  
15 invention may contain a single MAR as disclosed herein, multiple MARs of the present invention, or MARs of the present invention in conjunction with other MARs. A DNA construct of the present invention may comprise a first MAR of the present invention 5' to the transcription initiation region and a second MAR of a different sequence situated 3' to the structural gene, or vice versa.

20 The transcription initiation region, which preferably includes the RNA polymerase binding site (promoter), may be native to the host organism to be transformed or may be derived from an alternative source, where the region is functional in the host. Other sources include the Agrobacterium T-DNA genes, such as the transcriptional initiation regions for the biosynthesis of nopaline, octapine, mannopine,  
25 or other opine transcriptional initiation regions, transcriptional initiation regions from plants, transcriptional initiation regions from viruses (including host specific viruses), or partially or wholly synthetic transcription initiation regions. Transcriptional initiation and termination regions are well known. See, e.g., dGreve, *J. Mol. Appl. Genet.* 1, 499-511 (1983); Salomon et al., *EMBO J.* 3, 141-146 (1984); Garfinkel et al., *Cell* 27, 143-30 153 (1983); and Barker et al., *Plant Mol. Biol.* 2, 235-350 (1983).

The transcriptional initiation regions may, in addition to the RNA polymerase binding site, include regions which regulate transcription, where the

regulation involves, for example, chemical or physical repression or induction (e.g., regulation based on metabolites or light) or regulation based on cell differentiation (such as associated with leaves, roots, seed, or the like in plants). Thus, the transcriptional initiation region, or the regulatory portion of such region, is obtained from an appropriate gene which is so regulated. For example, the 1,5-ribulose biphosphate carboxylase gene is light-induced and may be used for transcriptional initiation. Other genes are known which are induced by stress, temperature, wounding, pathogen effects, etc.

Structural genes are those portions of genes which comprise a DNA segment coding for a protein, polypeptide, or portion thereof, possibly including a ribosome binding site and/or a translational start codon, but lacking a transcription initiation region. The term can also refer to introduced copies of a structural gene where that gene is also naturally found within the cell being transformed. The structural gene may encode a protein not normally found in the cell in which the gene is introduced or in combination with the transcription initiation region to which it is operationally associated, in which case it is termed a heterologous structural gene. Genes which may be operationally associated with a transcription initiation region of the present invention for expression in a plant species may be derived from a chromosomal gene, cDNA, a synthetic gene, or combinations thereof. Any structural gene may be employed. Where plant cells are transformed, the structural gene may encode an enzyme to introduce a desired trait, such as glyphosphate resistance; a protein such as a *Bacillus thuringiensis* protein (or fragment thereof) to impart insect resistance; or a plant virus protein or fragment thereof to impart virus resistance.

The expression cassette may be provided in a DNA construct which also has at least one replication system. For convenience, it is common to have a replication system functional in *Escherichia coli*, such as ColE1, pSC101, pACYC184, or the like. In this manner, at each stage after each manipulation, the resulting construct may be cloned, sequenced, and the correctness of the manipulation determined. In addition, or in place of the *E. coli* replication system, a broad host range replication system may be employed, such as the replication systems of the P-1 incompatibility plasmids, e.g., pRK290. In addition to the replication system, there will frequently be at least one marker present, which may be useful in one or more hosts, or different markers

for individual hosts. That is, one marker may be employed for selection in a prokaryotic host, while another marker may be employed for selection in a eukaryotic host, particularly a plant host. The markers may be protection against a biocide, such as antibiotics, toxins, heavy metals, or the like; provide complementation, for example by imparting prototrophy to an auxotrophic host; or provide a visible phenotype through the production of a novel compound. Exemplary genes which may be employed include neomycin phosphotransferase (NPTII), hygromycin phosphotransferase (HPT), chloramphenicol acetyltransferase (CAT), nitrilase, and the gentamicin resistance gene. For plant host selection, non-limiting examples of suitable markers are  $\beta$ -glucuronidase, providing indigo production, luciferase, providing visible light production, NPTII, providing kanamycin resistance or G418 resistance, HPT, providing hygromycin resistance, and the mutated aroA gene, providing glyphosate resistance.

The various fragments comprising the various constructs, expression cassettes, markers, and the like may be introduced consecutively by restriction enzyme cleavage of an appropriate replication system, and insertion of the particular construct or fragment into the available site. After ligation and cloning the DNA construct may be isolated for further manipulation. All of these techniques are amply exemplified in the literature and find particular exemplification in Sambrook et al., Molecular Cloning: A Laboratory Manual, (2d Ed. 1989)(Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Vectors which may be used to transform plant tissue with DNA constructs of the present invention include vectors used for *Agrobacterium*-mediated transformation and ballistic vectors, as well as vectors suitable for direct DNA-mediated transformation.

Microparticles carrying a DNA construct of the present invention, which microparticles are suitable for the ballistic transformation of a cell, are also useful for transforming cells according to the present invention. The microparticle is propelled into a cell to produce a transformed cell. Where the transformed cell is a plant cell, a plant may be regenerated from the transformed cell according to techniques known in the art. Any suitable ballistic cell transformation methodology and apparatus can be used in practicing the present invention. Exemplary apparatus and procedures are disclosed in Stomp et al., U.S. Patent No. 5,122,466; and Sanford and Wolf, U.S. Patent

No. 4,945,050 (the disclosures of all U.S. Patent references cited herein are incorporated herein by reference in their entirety). When using ballistic transformation procedures, the expression cassette may be incorporated into a plasmid capable of replicating in the cell to be transformed. Examples of microparticles suitable for use in such systems 5 include 1 to 5  $\mu\text{m}$  gold spheres. The DNA construct may be deposited on the microparticle by any suitable technique, such as by precipitation.

Plant species may be transformed with the DNA construct of the present invention by the DNA-mediated transformation of plant cell protoplasts and subsequent regeneration of the plant from the transformed protoplasts in accordance 10 with procedures well known in the art.

Any plant tissue capable of subsequent clonal propagation, whether by organogenesis or embryogenesis, may be transformed with a vector of the present invention. The term "organogenesis," as used herein, means a process by which shoots and roots are developed sequentially from meristematic centers; the term 15 "embryogenesis," as used herein, means a process by which shoots and roots develop together in a concerted fashion (not sequentially), whether from somatic cells or gametes. The particular tissue chosen will vary depending on the clonal propagation systems available for, and best suited to, the particular species being transformed. Exemplary tissue targets include leaf disks, pollen, embryos, cotyledons, hypocotyls, 20 megagametophytes, callus tissue, existing meristematic tissue (e.g., apical meristems, axillary buds, and root meristems), and induced meristem tissue (e.g., cotyledon meristem and hypocotyl meristem).

Plants of the present invention may take a variety of forms. The plants may be chimeras of transformed cells and non-transformed cells; the plants may be 25 clonal transformants (e.g., all cells transformed to contain the expression cassette); the plants may comprise grafts of transformed and untransformed tissues (e.g., a transformed root stock grafted to an untransformed scion in citrus species). The transformed plants may be propagated by a variety of means, such as by clonal propagation or classical breeding techniques. A dominant selectable marker (such as 30 *npt* II) can be associated with the expression cassette to assist in breeding.

Plants which may be employed in practicing the present invention include (but are not limited to) tobacco (*Nicotiana tabacum*), potato (*Solanum*

*tuberosum*), soybean (*glycine max*), peanuts (*Arachis hypogaea*), cotton (*Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (5 *Persea americana*), fig (*Ficus carica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), corn (*Zea mays*), wheat, oats, rye, barley, rice, vegetables, 10 ornamentals, and conifers. Vegetables include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Pisum* spp.) and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals 15 include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosa-sinensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum. Gymnosperms which may be employed to carrying out the present invention include conifers, including pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliottii*), ponderosa pine (*Pinus ponderosa*), 20 lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). 25

The examples which follow are set forth to illustrate the present invention, and are not to be construed as limiting thereof.

#### EXAMPLE 1

##### Materials and Methods

###### 30 *NT-1 Protoplast Isolation*

One hundred ml cultures of four day old tobacco NT-1 suspension cells were spun at 1400 rpm (585xg) for five minutes in a Beckman GPR table top centrifuge rotor

GH 3.7 and washed in 10 mM MES (2-[N-morpholino]ethane-sulfonic acid sodium salt, Sigma M-3885) pH 5.5, 0.4M mannitol. The pellet was resuspended in 100 ml of 10mM MES pH 5.5, 0.4M mannitol containing 1g of cellulase (Onozuka RS Yakult Pharmaceutical LTD) and 0.1g of pectolyase (Y-23 Seishin Corp.) and incubated for 30  
5 to 60 minutes at 28° C with gentle shaking in order to remove the cell wall. The resulting protoplasts were pelleted at 1400 rpm and washed two times in 50ml of cold (4°C) 0.4 mannitol (unbuffered).

#### *NT-1 Nuclei Isolation*

10 The protoplasts were pelleted and resuspended in 50ml of Nuclei Isolation Buffer 1 (NIB1) at pH 6.5 (NIB1 = 0.5M hexylene glycol, 20mM N-2-hydroxyethylpiperazine-N-ethanesulfonic acid (hepes), 20mM KCl, 1% thiodiglycol, 50mM spermine (Sigma S-2876), 125 mM spermidine (Sigma S-2501), 0.5mM phenylmethylsulfonyl fluoride (PMSF 2M stock in methanol), 2µg/ml aprotinin (Sigma  
15 A-6279), 0.5% Triton X-100, 0.5mM EDTA). This procedure solubilizes the plasma membrane and releases nuclei. After a five minute incubation on ice, the nuclei were filtered through a tier of 100um, 50um, and 30um nylon mesh to remove the cellular debris and then spun through 15% Percoll (Pharmacia 17-0891-01)/NIB1 for further purification. The pelleted nuclei were washed two times with of Nuclei Isolation Buffer  
20 2 (NIB2 = NIB1 without EDTA).

#### *Quantification of Nuclei*

The nuclei were resuspended in a suitable volume of storage buffer (NIB2 in 50% glycerol) such that the suspension would have an absorbance reading of 10 at 260  
25 nm. Absorbance was determined by diluting 2 µl of nuclei in 0.5 ml of 2.2M sodium chloride, 5.5M urea. One ml aliquots were stored at -70°C until needed. The number of nuclei per tube was determined by counting aliquots using a hemocytometer. Although there was some variation between preparations, in general, each 1ml tube with an absorbance of 10 contains about 3.5 million nuclei.

30

#### *Preparation of Nuclear Halos and Nuclear Matrices*

Approximately 3.5 million nuclei (one tube stored at -70°C) were thawed on ice

and washed in 10ml of Nuclei Isolation Buffer 3 (NIB3 = 0.5M hexylene glycol, 20mM hepes pH 7.4, 20mM KCl, 1% thioglycol, 50mM spermine, 125 mM spermidine, 0.5mM PMSF, 2 $\mu$ g/ml aprotinin). The nuclei were pelleted at 1400 rpm, resuspended in 200  $\mu$ l of NIB3 containing 1mM CuSO<sub>4</sub> and incubated at 42°C for 15 minutes in  
5 order to stabilize the nuclear matrix.

To remove the histones and other soluble proteins, the nuclei were incubated in 10ml of Halo Isolation Buffer 2 (HIB2 = 10mM 3,5 diiodosalicylic acid lithium salt (Sigma D-3635), 100mM lithium acetate, 20mM hepes, 2mM EDTA, 0.1% digitonin, 0.5mM PMSF, 2 $\mu$ g/ml aprotinin) for 15 minutes at room temperature. (Digitonin is  
10 prepared by mixing 5g in 12.5 ml of methanol, heating to 65°C to dissolve, filtering through Whatman #1 filter paper, recrystallizing by removing the methanol under vacuum, weighing the resulting crystals and resuspending in water at a concentration of 5%, and storing at -20°C until needed). When histones are removed the coiling restraints on the DNA are removed, allowing the DNA to spill out of the nucleus to form  
15 a 'nuclear halo'.

The nuclear halos were pelleted at 3600 rpm (2900xg) and then washed with 10ml of Digestion/Binding Buffer (D/BB = 70mM NaCl, 20mM Tris-HCl pH 8.0, 20mM KCl, 0.1% digitonin, 1% thioglycol, 50mM spermine, 125mM spermidine, 2 $\mu$ g/ml aprotinin and 0.5mM PMSF). The second wash contained all the elements of  
20 the first wash, plus 10uM phenanthroline, and the third wash contained all the elements of the second was plus 10mM MgCl<sub>2</sub>. The halos were resuspended in 500 $\mu$ l of D/BB plus all the elements of wash three.

For a discussion of the above-described methods, see Hall and Spiker, *Plant Molecular Biology Manual* D2: 1-12, Kluwer Academic Publishers, Dordrecht, the  
25 Netherlands.

To cleave the DNA, nuclear halos were treated with nucleases (either 250U of RsaI, TaqI, EcoRI, HindIII, or with DnaseI (Sigma) at 0.1 $\mu$ g/ml) and incubated at 37°C for 90 minutes with the addition of another 250U of restriction enzymes or 0.1 $\mu$ g/ml of DnaseI after 45 minutes. The resulting nuclear matrices and their associated DNA were  
30 separated from unbound DNA by centrifugation at 3600 rpm for 5 minutes.

*Isolation of functionally-defined MAR DNA*

Nuclear matrices were washed with 1ml of D/BB and 10mM MgCl<sub>2</sub> to remove residual supernatant (unbound) DNA and to remove protease inhibitors. The pellet was 5 resuspended in 500 µl of protease buffer (10mM Tris-HCl pH 8.0, 20mM EDTA, 0.5% SDS, 0.5mg/ml proteinase K) and incubated at room temperature overnight. The matrix bound DNA was further purified by phenol:chloroform extraction and ethanol precipitation, dried and resuspended in 100 µl of Tris-EDTA (TE = 10mM Tris-HCl pH8.0 and 10mM EDTA).

10

*Cloning*

The purified operationally defined MAR DNA fragments were cloned into pBluescript II SK+ (Stratagene). The vector was digested with either EcoRV, Clal, EcoRI, Hind III or HincII (for blunt end ligation of DnaseI generated fragments), and 15 ligated (using New England Biolabs T4 ligase according to the manufacturer's protocol) to the purified DNA from the nuclear matrices purified with RsaI, TaqI, EcoRI, HindIII, or DnaseI. Stratagene *E. coli* SURE cells were transformed with the plasmids according to the manufacturer's protocol.

20 *Isolation of Plasmid DNA*

Plasmid DNA was isolated from transformants that were grown in 2 ml of Luria Broth (10g/l tryptone, 10g/l yeast extract, 5g/l NaCl) with 80 µg/ml ampicillin overnight at 37°C with shaking. The cells were spun at 13000 rpm in a microfuge for 2 minutes, and the pellets were resuspended in 150µl of 20% sucrose, 25mM Tris-HCl pH 8.0, 25 10mM EDTA. The cells were treated with 350µl of lysis buffer (1% SDS and 200 mM NaOH) and incubated at room temperature for 10 minutes. After the addition of 250 µl of 3M sodium acetate pH 5.2, the cells were incubated on ice for 10 minutes and then spun at 13000 rpm for 20 minutes at 4°C. The plasmid-DNA containing supernatant was transferred to a fresh tube containing 0.7ml of isopropanol and spun at 13000 rpm 30 for 20 minutes at 4°C. The pellets were washed with 70% ethanol, air dried overnight and then resuspended in 50µl of TE containing 5µg of Rnase A (Sigma), incubated at 37°C for 1 hour and stored at 4°C until needed. Alternatively, when large quantities of

DNA were required, plasmid DNA was isolated using Qiagen columns according to the manufacturer's protocol.

*End Labeling Protocol*

5       Plasmid DNAs isolated from individual transformants were end labeled and tested for binding to the nuclear matrix. One µg of plasmid DNA was digested with the appropriate enzymes to release the fragment (usually EcoRI and HindIII) according to the manufacturer's protocol. A standard end-labeling reaction contained 250ng of digested DNA (5µl), 0.5µl of 10X Klenow buffer, 0.33µl of dNTPs (2mM deoxycytosine triphosphate, 2mM deoxyguanidine triphosphate and 2mM deoxythymidine triphosphate), 2.5 µl of 10mCi/ml  $\alpha$ -<sup>32</sup>P deoxyadenosine triphosphate (Dupont NEN BLU-012H) and 0.2µl of 5,000U/ml DNA polymerase large fragment (NEB Klenow) in a total of 10µl. The mixture was incubated at room temperature for 15 minutes and the reaction stopped by the addition of 40µl of TE. The unincorporated nucleotides were removed by centrifugation through a Sephadex G-50 spin column. The amount of radioactivity (counts per minute) in the resulting end-labeled DNA was determined by placing 2µl of labeled DNA in 3 ml of Scinti Verse (Fisher SX 1-4) and counted using the Beckman LS 100C scintillation counter.

20      *Matrix Binding-Exogenous Assay*

Nuclear halos were treated with 250U of EcoRI and HindIII for 90 minutes at 37°C with the addition of another 250U of each enzyme after 45 minutes. The resulting nuclear matrices were aliquoted at 50 µl for different binding reactions (different labeled DNA fragments for testing). Each 50µl aliquot contained one tenth of the nuclear matrices, about 350,000, as well as one tenth of the cleaved, non-MAR, endogenous DNA, which served as nonspecific competitor. Radioactively labeled DNA fragments of interest were incubated with the nuclear matrices at 50,000 cpm per fragment (about 5ng of DNA) per 50µl reaction at 37°C for 3 hours with resuspension every 20 minutes. The pellet and supernatant fractions were separated by centrifugation at 3600 rpm for 5 minutes. The supernatant was transferred to a fresh tube containing 0.5µl of 0.5M EDTA pH 8.0 and stored at -20°C. The matrices were washed with 200 µl of D/BB plus 10mM MgCl<sub>2</sub> (D/BB = 70mM NaCl, 20mM Tris-HCl pH 8.0, 20mM KCl, 0.1%

digitonin, 1% thiodiglycol, 50mM spermine, 125mM spermidine) and then resuspended in 50 $\mu$ l of protease buffer (10mM Tris-HCl pH 8.0, 20mM EDTA, 0.5% SDS, 0.5mg/ml proteinase K) and incubated at room temperature overnight. Twenty  $\mu$ l aliquots of the pellet and supernatant fractions were subjected to electrophoresis in a 1% agarose (FMC Sea Kem GTG 50072) gel, prepared and run in TAE (TAE = 40mM Tris-acetate pH 8.0 and 1mM EDTA). The gel was then treated for 20 minutes in 7% trichloroacetic acid, dried and exposed to X-ray film (Kodak X-OMAT AR). This method of representing the DNA bound to the nuclear matrix is called the 'equal fractions' method, as an equal portion of the DNA from the pellet and the supernatant fractions (e.g., 20%) is applied to the gel. This approach allows direct determination of the amount of a fragment partitioning with the pellet or supernatant; very weak-binding DNA fragments are not scored as MARs.

## EXAMPLE 2

15        **Results of Isolation and Testing of Operationally Defined MARs**

A random sample of MAR fragments was obtained by purifying matrix associated DNA and cloning these fragments. Five different preparations of nuclear matrices were made using one of five different nucleases. From each preparation, twenty colonies were picked, grown and analyzed for the presence of single inserts. A 20 total of thirty-nine clones were then tested for their ability to rebind to the nuclear matrix. Since all clones were obtained using one of the five cloning strategies, the plasmids obtained using each strategy differ only in the content of their insert. The generalized plasmid maps are summarized in **Figure 1**.

Each clone was end labeled and tested for the ability to rebind to the nuclear 25 matrix using the exogenous assay. A previously identified strong binding MAR fragment, ToRB7-6 (Hall et al., *Proc. Natl. Acad. Sci. USA* 88:9320 (1991)), served as a positive control. In each case, the non-binding vector served as an internal negative control. Results of such binding assays are shown in **Table 1**, which contains a summary of the relative binding strength of all the MAR clones tested. Among the 30 clones obtained from nuclear halos treated with restriction enzymes that have four base pair recognition sites (*Rsa*I and *Taq*I), 9 of 17 fragments had some binding activity as compared to 14 of 17 clones when enzymes with six base pair recognition sites were

used. In addition to the 34 clones represented, five clones from the DnaseI treated nuclear halos were tested (302, 303, 305, 311 and 319); no binding was detected for any of these samples (results not shown). Relative binding strength was based on the proportion of each MAR fragment that partitioned in the bound fraction on a scale of 0 to 100%. In Table 1, no = no detectable binding; weak = detectable -40%; medium = 40-70%; strong = 70-100%.

**Table 1**  
**Clones Tested for Matrix Binding Activity\***

Four base cutters		Six base cutters	
Clone #	Binding strength	Clone #	Binding strength
RsaI : EcoRV		HindIII	
1	Weak	201	weak
2	Weak	202	medium
3	No	203	no
4	Weak	205	medium
6	no	206	weak
7	weak	209	weak
8	no	211	strong
21	no	216	weak
22	weak	210	no
23	no	217	weak
26	no	218	no
27	no	219	weak
28	weak	220	medium
TaqI : ClaI		EcoRI	
11	No	109	weak
15	Weak	113	weak
34	Weak	115	medium
35	Weak	116	strong

\* Relative binding strength is based on the proportion of each MAR fragment that partitioned into the bound fraction on a scale of 0 to 100%, where no = no detectable binding; weak = detectable - 40% binding; medium = 40% - 70% binding; and strong = 70% - 100% binding.

Because the fragments were isolated through their association with the nuclear matrix (the operational definition of a MAR), all of the fragments would be expected to rebind to the matrix in the exogenous assay. However, 40% of the clones did not have detectable binding activity. There are several possible explanations for this discrepancy. One such possibility is a cloning artifact. Some of the DNA fragments may have been altered during cloning, resulting in a loss of binding activity. This possibility can be

substantiated in at least one case, pS8. In this clone the expected sequence at the ligation site is GATAC, but sequencing revealed GATCA. This indicates that the fragment was altered during the procedure. Since the clone was created by blunt end ligation, it is likely that the fragment was broken during some procedure and quite 5 possible that the resulting fragment had lost its binding capability. Another possibility is that some of the non-MAR DNA was trapped within the nuclear matrix during isolation. This is a rare phenomenon, and would not be expected to result in 40% non-MAR clones. It is also possible that some of the non-binding clones really are bound to the nuclear matrix *in vivo* but that the sensitivity of the *in vitro* exogenous assay was not 10 sufficient to detect weak binding fragments (see Materials and Methods for discussion of sensitivity of detection). Another possible explanation for the isolation of operationally defined MAR sequences that did not rebind to the nuclear matrix is their size. Several of the isolated MAR fragments showed significant reduction in binding 15 when cleaved into smaller pieces (data not shown), which supports the hypothesis of a lower size limit for MAR fragments. The clones ranged in size from 150 to 200 base pairs. It is possible that none of the isolated DnaseI clones had binding activity because they were below the minimum size requirement. Furthermore, a higher percentage of 20 the clones obtained using restriction enzymes with six base pair recognition sites (EcoRI and HindIII) had matrix binding activity (14/17) than those clones obtained using restriction enzymes with four base pair recognition sites (9/17). Again, this subtle discrepancy may be the result of the smaller size of the fragments generated using four 25 base cutters. This observation may also play a part in the strength of binding, since all of the clones isolated using restriction enzymes with four base pair recognition sites were weak binding MARs or had no detectable binding. It should be noted that although there appears to be a lower limit of size, within a population of fragments of 30 the appropriate size (300 bp to several kilobases), there does not appear to be a statistically significant correlation between binding strength and the size of the MAR fragment (see Figure 5). Finally, some of the non-binding clones may be parts of MARs that were cleaved during isolation. It appears that the binding of MAR DNA to the nuclear matrix may involve extended contact. If a particular sequence is cut within the extended contact region, it is possible that none of the resulting fragments would be able to rebind to the nuclear matrix in the exogenous assay.

**EXAMPLE 3****Materials and Methods for Sequence Analysis of MARs***Subcloning and Sequencing*

5 To effectively analyze the sequence of the MAR DNA, DNA that is not necessary for binding must be excluded. Core binding fragments of the MARs were identified by testing the binding of different subfragments of the clones and determining the particular fragments that contained the majority of the binding activity. These fragments were then cloned and sequenced.

10 A 998 bp fragment of pS116 was excised using MunI and PstI and ligated into the EcoRI and PstI sites of pBluescript II SK+ to create pS116-1.1B (SEQ ID NO:5).

A 635 bp and a 1087 bp fragment of pS202 were excised by cleavage with HindIII and EcoRI and ligated into pBC KS+ (Stratagene) using the same two restriction enzymes to create pS202-1 (SEQ ID NO:6) and pS202-2 (SEQ ID NO:7),  
15 respectively.

A 704bp fragment of pS205 was excised with EcoRI and HindIII and ligated into the EcoRI and HindIII sites of pBC KS+ to create pS205-2 (SEQ ID NO:8).

A 306 bp fragment of pS206 was excised with BamHI and HindIII and ligated into the BamHI and HindIII sites of pBC KS+ to create pS206-1 (SEQ ID NO:9).  
20

A 685bp fragment of pS211 was excised with EcoRI and HindIII and ligated into the EcoRI and HindIII sites of pBC KS+ to create pS211-1 (SEQ ID NO:10).

A 899bp fragment of pS217 was excised with BamHI and ligated into the BamHI site of pBC KS+ to create pS217-1 (SEQ ID NO:11).

25 A 1499 bp fragment of pS220 was excised with XhoI and HindIII and ligated into the XhoI and HindIII sites of pBC KS+ to create pS220-1 (SEQ ID NO:13).

All digestions, ligations and transformations were performed according to manufacturer protocols. The maps of the original plasmids and their subclones are depicted in **Figure 2**.

The MAR fragments were sequenced in stages by primer walking. Each clone was sequenced using the Universal-21M13 (TGTAAAACGA CGGCCAGT)(SEQ ID NO:18) and reverse M13 (CAGGAAACCGA TATGACC)(SEQ ID NO:19) primers at Iowa State University Nucleic Acids Facility for the initial portions of the sequence. In 5 the case of longer clones, when a complete sequence was not obtained from the initial sequence, internal primers were designed and then constructed at the Molecular Genetics Facility at North Carolina State University. All interior primers are underlined in **Figure 3**. Because of the AT-rich nature of the MARs, primers with suitable melting temperatures, but that avoided secondary structural features, had to be designed. The 10 usefulness of each primer was tested by attempting to amplify an internal MAR fragment using the constructed primer in conjunction with either the universal or reverse primer in a polymerase chain reaction (PCR). PCR was performed using Boehringer Mannheim Taq polymerase according to the manufacturer's protocol. Primers that resulted in successful amplification of DNA were used for additional sequencing at Iowa 15 State University. Sequences of the primers used are underlined within the sequences in **Figure 3**.

#### EXAMPLE 4

##### MAR Sequences

20 From the randomly isolated MAR fragments, a sample of ten binding clones (pS1, pS4, pS115, pS116, pS202, pS205, pS206, pS211, pS217 and pS220) and two non-binding clones (pS8 and pS218) were chosen for sequence analysis. These particular MARs were chosen as representatives of the population based on binding strength. Within this population of ten sequences are several weak, medium and strong 25 binding MARs, representing the spectrum of binding strengths within population. The binding of MAR fragments appears to involve multiple protein DNA interactions rather than a single binding site (Gasser et al., 1989), however, it is not known if multiple interactions are required for MAR binding, or if longer MAR fragments simply consist of many smaller MARs each acting independently. The binding of a particular MAR 30 may be affected when that fragment is cleaved so that shorter fragments may contain none, some or all of the binding potential of the original full-length clone. In some

instances, all of the fragments are capable of binding to the nuclear matrix, whereas in others the binding is confined to one of the smaller fragments.

Several of the isolated clones were several kilobases in length. To avoid including non-MAR DNA in our sequence analysis, these samples were digested into 5 smaller fragments and the binding of these subfragments tested. In the case of seven clones (116, 202, 205, 206, 211, 217 and 220) a smaller core binding fragment that maintained the most of the original binding strength was identified. These core binding fragments were subcloned and used instead of the original sequence (see **Figure 2** for plasmid maps). In the case of clone pS202, two binding subfragments (pS202-1 (SEQ 10 ID NO:6) and pS202-2 (SEQ ID NO:7)) were identified, each of which maintained a binding strength similar to that of the original clone. Both fragments are included in this study. The other three clones (pS1, pS4, and pS115) were sequenced in their entirety (SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:4, respectively). Two of the three non-binding clones (pS8 and pS218) were sequenced in their entirety (SEQ ID NO:3 and 15 SEQ ID NO:12, respectively). The binding strengths of the subclones were rated on a scale of 0-100 based on the percent of the MAR that partitioned in the bound fraction in the standard exogenous assay (**Figure 4**).

## EXAMPLE 5

### Motif Significance in MAR Sequences

Several different motifs have previously been identified as associated with MAR sequences, including A boxes (SEQ ID NO:14), T boxes (SEQ ID NO:15), ARS consensus (SEQ ID NO:16), and the Drosophila consensus sequence for topoisomerase II (SEQ ID NO:17). A search was conducted for the presence of these motifs using the 20 Apple MacIntosh program MacVector. In the search, 1 mismatch was allowed in the cases of the ARS consensus and T box, and 2 mismatches in the cases of A box and the consensus for topoisomerase II. Allowing for this number of mismatches results in 25 similar probabilities of occurrence for all four motifs (see **Table 2**). The A box and T box motifs often resulted in overlapping regions due to their AT-rich nature. For example, within a stretch of 20 bases of Ts, 10 T boxes can be found. The probability of 30 finding an additional T box upon inclusion of the next base is 0.35 in a region of DNA

that contains 70% AT. To avoid this artificial identification of additional motifs, only two T boxes would be counted in the above example.

In the MAR DNA fragments obtained from the random cloning procedures described in the above Examples, AT rich motifs were present. These AT rich regions 5 are depicted in Figure 4. Additionally, several of the randomly obtained MAR sequences contained short stretches (20bp) of highly AT rich ( $\geq 90\%$ ) DNA. The locations of these stretches are also shown in Figure 4.

The present inventors tested the occurrence of each of these motifs for statistical significance against the number of occurrences that would be expected at random, to 10 determine if these motifs are over-represented in MAR DNA. The probability of the presence of a specific sequence starting at a specific base within 70% AT rich DNA was calculated by multiplying the probabilities of each base in the motif (0.35 for A and T, 0.15 for G and C). Since mismatches of 1 or 2 bases were allowed for certain motifs, the probabilities were adjusted by dividing the calculated probability of occurrence of 15 the motif with no mismatches by the lowest probability of an individual base (or bases when two mismatches were allowed) within the motif, yielding a conservative estimate of probability. In addition, the probabilities were multiplied by a factor of two, since these motifs can occur in either strand of the DNA. Although the two strands of a DNA sequence are not independent, the factor of two provides a simple and conservative 20 method for calculating the expected frequency within double stranded DNA. For example, the probability of the topoisomerase II consensus sequence (GTNWAYATTNATNNR) occurring in 70% AT rich DNA without any mismatches is calculated by multiplying the expected frequency of each base together:

$$[(0.15)(0.35)(1)(0.7)(0.35)(0.5)(0.35)(0.35)(0.35)(0.35)(1)(0.35)(0.35)(1)(1)(0.5)] = \\ 25 \quad 1.68 \times 10^{-5}.$$

This calculated probability is then adjusted for the allowance of up to two mismatches by dividing by 0.15 and 0.35, the two lowest expected frequencies of each base. The resulting value is then multiplied by two, since motifs can occur in either strand of DNA. The calculated probability for an A box with two mismatches, an ARS 30 consensus with 1 mismatch, and the consensus for topoisomerase II with two mismatches are coincidentally identical,  $6.428 \times 10^{-4}$ , whereas the calculated probability of a T box occurring with one mismatch is slightly higher,  $1.26 \times 10^{-3}$ . The probability

of the occurrence of a 20 bp stretch of DNA containing 90% or greater AT content, was calculated by taking the probability of a single base being either A or T (0.7) to the 18<sup>th</sup> power. Therefore, the probability of occurrence of a twenty base pair stretch with two mismatches (or 90% AT rich DNA) is equal to  $1.628 \times 10^{-3}$ . Since this type of motif 5 will automatically be present on both strands at the same time, only those motifs found in one strand were counted. As with the other motifs, overlapping regions were not counted as separate occurrences, however, this motif often occurs in stretches much longer than 20 bp. The actual lengths of the regions are depicted in Figure 4. The number of occurrences were counted as one from 20-39 bp; two from 40-59 bp; and so 10 on. The probability of a motif starting at any one site, was converted to an expected number of occurrences within a DNA sequence by multiplying by the number of base pairs in its length.

For purposes of the calculations the probability of a particular base occurring at a particular site was assumed to follow a multinomial distribution, i.e., any of the four bases can occur at a particular site. Note that this assumption may not be true if there 15 are restraints on DNA sequences (e.g., five consecutive Gs not allowed), but since such restraints are presently unknown, the assumption of multinomial distribution was made. It would be expected that the probability of a particular string of bases (a motif) occurring at a particular site to have approximately a normal distribution. The observed 20 number of MAR motifs were compared to the number expected under this normal distribution assumption. The number of observed MAR motifs was significantly greater (at the 1% level) than the number expected if the observed number of MAR motifs is greater than a calculated critical value. The critical value is  $\mu + Z_{0.01}\sigma$  (*Steele and Torrie, Principles and Procedures of Statistics: A Biometrical Approach*, McGraw-Hill Publishing Co., New York, NY (1980)). In this equation  $\mu$  is the expected number of 25 MAR motifs,  $Z_{0.01}$  is the Z statistic at 1% and  $\sigma$  is the standard deviation of the expected number of occurrences of MAR motifs given a random expected number of occurrences given a random sequence of bases ( $\sigma = \sqrt{\mu}$ ). The motifs in question have already been shown to be associated with certain MAR sequences. The null hypothesis of the motifs occurring at random is compared to the one tailed test of the alternative hypothesis, that 30 these motifs occur more often than would be expected by random occurrence. The 1% critical values are shown in Table 2. Since the observed number of occurrences for all

five motifs is greater than the critical values, the null hypothesis can be rejected in favor of the alternative hypothesis that these motifs occur more often in MAR DNA than would be expected by chance alone. A boxes, T boxes, ARS consensus, topoisomerase II consensus and 20 bp regions of  $\geq$  90% AT DNA all occur more often than would be expected at random.

Clone	length bp	<u><math>\Delta</math>box-2</u> <u><math>6.428 \times 10^{-4}</math></u>		<u><math>\Delta</math>ARS con-1</u> <u><math>6.428 \times 10^{-4}</math></u>		<u><math>\Gamma</math>box-1</u> <u><math>1.261 \times 10^{-3}</math></u>		<u><math>\Gamma</math>box II-2</u> <u><math>6.428 \times 10^{-4}</math></u>		<u><math>\Delta</math>AT</u> <u><math>1.28 \times 10^{-3}</math></u>	
		obs	exp	obs	exp	obs	exp	obs	exp	obs	exp
pSI	437	9	0.280	0	0.280	3	0.551	1	0.280	7	0.711
pS4	587	7	0.377	2	0.377	4	0.740	2	0.377	4	0.955
pS15	866	4	0.556	2	0.556	5	1.092	0	0.556	4	1.409
pS16-1.1B	998	11	0.642	1	0.642	14	1.258	2	0.642	16	1.624
pS202-1	635	7	0.408	0	0.408	4	0.800	1	0.408	3	1.033
pS202-2	1087	7	0.698	3	0.698	4	1.370	2	0.698	3	1.769
pS205-2	704	9	0.452	4	0.452	5	0.887	2	0.452	9	1.146
pS206-1	306	5	0.196	1	0.196	2	0.385	1	0.196	5	0.498
pS211-1	685	7	0.440	2	0.440	11	0.863	4	0.440	9	1.115
pS217-1	899	10	0.578	2	0.578	5	1.133	1	0.578	2	1.463
pS220-1	1499	21	0.963	5	0.963	10	1.890	2	0.963	20	2.440
Total	8705	97	5.599	22	5.599	67	10.974	18	5.599	82	14.171
1% Critical Value		11.10		11.10		18.67		11.10		22.92	
<u>Other MARS</u>											
<u>ARS-1</u>	838	5	0.538	2	0.538	2	1.056	0	0.538	1	1.364
<u>TsRB7-6</u>	1103	13	0.709	3	0.709	7	1.390	3	0.709	16	1.795
<u>Sca/Cla</u>											
<u>Non-binding clones</u>											
pS8	383	2	0.246	0	0.246	1	0.482	2	0.246	2	0.623
pS218	999	5	0.642	2	0.642	5	1.259	4	0.642	0	1.626

## EXAMPLE 5

### Motif Frequency and Binding Strength

To determine if a correlation exists between the number of these motifs present  
5 within a particular MAR sequence and its binding strength, that data (Figures 6-11) was plotted and tested the significance of the regression coefficient, r, using an F test (Steele and Torrie, 1980). Included were all eleven binding sequences (SEQ ID NOs: 1, 2, 4-11, and 13); two non-binding sequences (SEQ ID NO:3 and SEQ ID NO:12); and two well characterized MARs (ToRB7-6 and ARS-1).

10 The correlation between binding strength and the length of the MAR as well as the %AT content of MARs was analyzed. The analysis (Table 3), shows a significant correlation between the number of 20 bp stretches of 90% or greater AT and the binding strength of MARs. The number of T boxes is also significant, as is the number of A boxes. No significant correlation between binding strength and the length of the MAR  
15 fragment nor the presence of any of the other MAR related motifs was detected in this analysis. However, there was a significant relationship between binding strength and overall AT content. This is the first report of a correlation between the abundance of certain MAR related motifs and MAR binding strength. In addition, the newly identified MAR related motif of local AT rich regions ( $\geq 90\%$ ), has a higher correlation  
20 to MAR binding strength than any of the previously identified motifs.

TABLE 3

Motif	R	F calc	significance
%AT	0.77	18.93	**
length	0.38	2.19	ns
#A box/kbp	0.19	0.49	ns
total #A box	0.57	6.48	*
# ARS/kbp	0.29	1.27	ns
total # ARS box	0.35	1.85	ns
#T box/kbp	0.64	9.45	**
total # T box	0.76	17.98	**
# TopoII/kbp	0.06	0.04	ns
total # TopoII	0.23	0.07	ns
# 90% AT/kbp	0.69	12.07	**
total # 90% AT	0.80	24.62	**

5                   $r = \text{regression coefficient} = \frac{[\sum XY - (\sum X \sum Y)/n]}{(\sum X^2 - [(\sum X)^2/n]) (\sum Y^2 - [(\sum Y)^2/n])}$

Fcalc = F statistic =  $(n-2)[r^2/(1-r^2)]$

ns = not significant

\* = significant at 95%

10                \*\* = significant at 99%

15                The relationship between certain MAR related motifs and MAR binding strength suggests that these motifs are general components of MARs. The lack of this relationship for other motifs suggests that these sequences are over-represented without being associated with MAR function, or may be related to only certain categories of MARs. In addition several other MAR related motifs (such as the asymmetric GA-rich stretches, or homopurine stretches) are found within some, but not all of the randomly

obtained sequences. The presence of such a motif may indicate a specific class of MARs, but since there is no information about the location of these random sequences within the genome, it cannot be inferred from the presence of this motif that a sequence is a MAR with a specific function (or even a MAR at all).

5 MARs may interact with the nuclear matrix through a variety of secondary structure motifs including a narrow minor groove, transiently single stranded regions and bent DNA. These structural motifs are expected to be present in the MARs described here because of their high AT content. A narrow minor groove is a feature of DNA containing long A tracts, a predicted feature within high AT content DNA.

10

#### EXAMPLE 5

##### Increasing Average Expression Levels

15 Earlier studies (Allen et al., *Plant Cell* 5:603 (1993)) showed that flanking a *GUS* reporter gene with two copies of a yeast MAR element (ARS-1) increased average GUS expression by 12-fold in stably transformed cell lines. In the present Example, the same cell line is transformed with constructs similar to those of Allen et al., 1993, but using a MAR having a sequence selected from SEQ ID NOs:1, 2, 4-11 and 13.

20 Transformation is achieved by mixing the appropriate reporter test plasmid and a selection plasmid, co-precipitating them onto microprojectiles, and bombarding plates of tobacco suspension culture cells as described previously (Allen et al., 1993). Antibiotic-resistant microcalli are selected and each callus is used to start an independent suspension culture cell line. Histochemical staining of segments from the original microcalli show that the staining intensity is greater in cell lines transformed with MAR 25 plasmids. After several weeks of growth, with weekly transfers, suspension cells are harvested, DNA is extracted from each cell line for Southern analysis and quantitative PCR assays, and portions of the same cell population are used to measure extractable reporter activity and expressed protein levels. Transgene copy number estimates and expression data are calculated. Levels of *GUS* gene expression, measured as *GUS* 30 enzyme activity, are assessed and compared to controls.

The foregoing examples are illustrative of the present invention, and are not to be

construed as limiting thereof. The invention is described by the following claims, with equivalents of the claims to be included therein.

**THAT WHICH IS CLAIMED IS:**

1. An isolated DNA molecule having a nucleotide sequence selected from the group consisting of:

5 (a) SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:13; and

10 (b) sequences that hybridize to isolated DNA of (a) above under conditions represented by a wash stringency of 0.3M NaCl, 0.03M sodium citrate, and 0.1% SDS at 60°C, and which encode a matrix attachment region.

2. A DNA construct comprising:

15 (a) a transcription initiation region and a structural gene positioned downstream from said transcription initiation region and operatively associated therewith; and

(b) a matrix attachment region according to claim 1 positioned either 5' to said transcription initiation region or 3' to said structural gene.

20 3. A DNA construct according to claim 2, wherein said matrix attachment region is 5' to said transcription initiation region.

4. A DNA construct according to claim 2, wherein said matrix attachment region is 3' to said structural gene.

25 5. A DNA construct according to claim 2, further comprising a second matrix attachment region that differs in sequence from said matrix attachment region according to claim 1.

6. A DNA construct comprising:

30 (a) a transcription initiation region and a structural gene positioned downstream from said transcription initiation region and operatively associated therewith;

(b) a matrix attachment region according to claim 1 positioned either 5' to said transcription initiation region or 3' to said structural gene; and

5 (c) a second matrix attachment region according to Claim 1, wherein said second matrix attachment region is positioned either 5' to said transcription initiation region or 3' to said structural gene.

7. A DNA construct according to claim 2, further comprising a termination sequence positioned downstream from said structural gene and operatively associated therewith.

10

8. A DNA construct according to claim 2, wherein said first and said second matrix attachment regions differ in sequence.

9. A vector comprising a DNA construct according to claim 2.

15

10. A vector according to claim 9, wherein said vector is selected from the group consisting of plasmids, viruses, and plant transformation vectors.

20

11. A host cell containing a DNA construct according to claim 2.

12. A host cell according to claim 9, wherein said host cell is an animal cell or a plant cell.

25 13. A transgenic plant comprising transformed plant cells, said transformed plant cells containing a DNA construct according to claim 2.

14. A transgenic plant according to claim 13, which is a monocot.

15. A transgenic plant according to claim 13, which is a dicot.

30

16. A transgenic plant according to claim 13, which plant is a dicot selected from the group consisting of tobacco, potato, soybean, peanuts, cotton, and

vegetable crops.

17. A DNA construct comprising, in the 5' to 3' direction, a transcription initiation region, a structural gene positioned downstream from said transcription initiation region and operatively associated therewith, and a matrix attachment region positioned either 5' to said transcription initiation region or 3' to said structural gene, wherein said matrix attachment region has a sequence selected from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:13;

said DNA construct carried by a plant transformation vector.

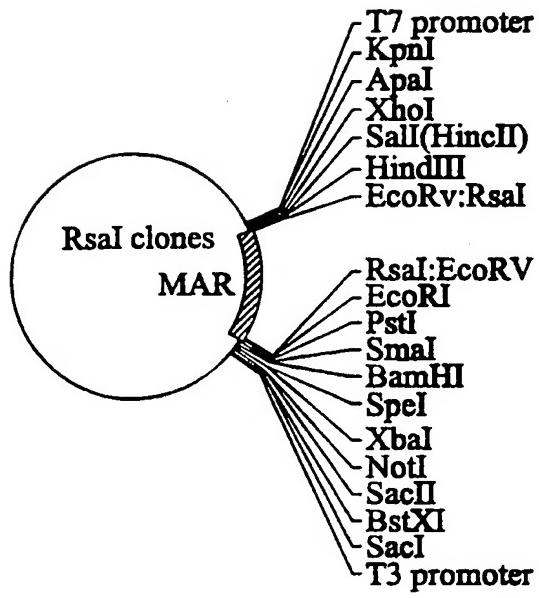
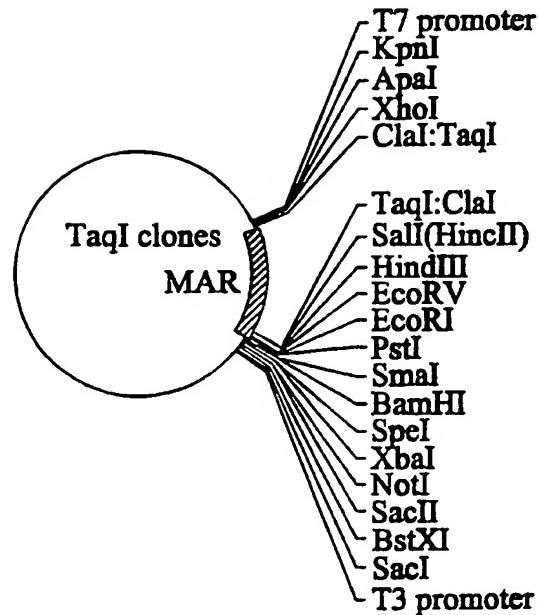
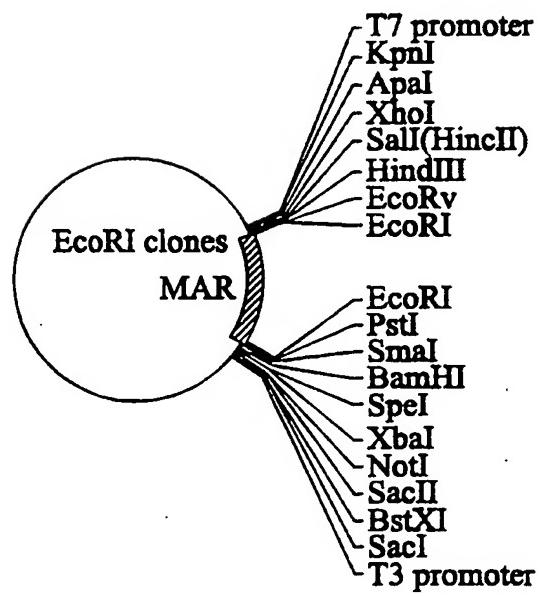
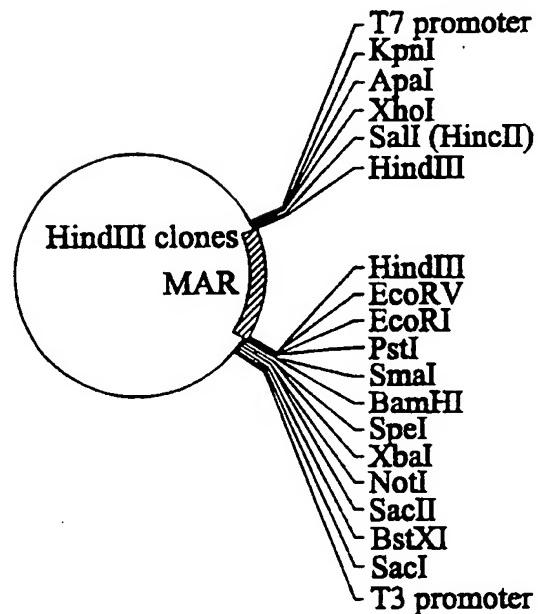
18. A DNA construct according to claim 17, further comprising a second matrix attachment region that differs in sequence from said matrix attachment region.

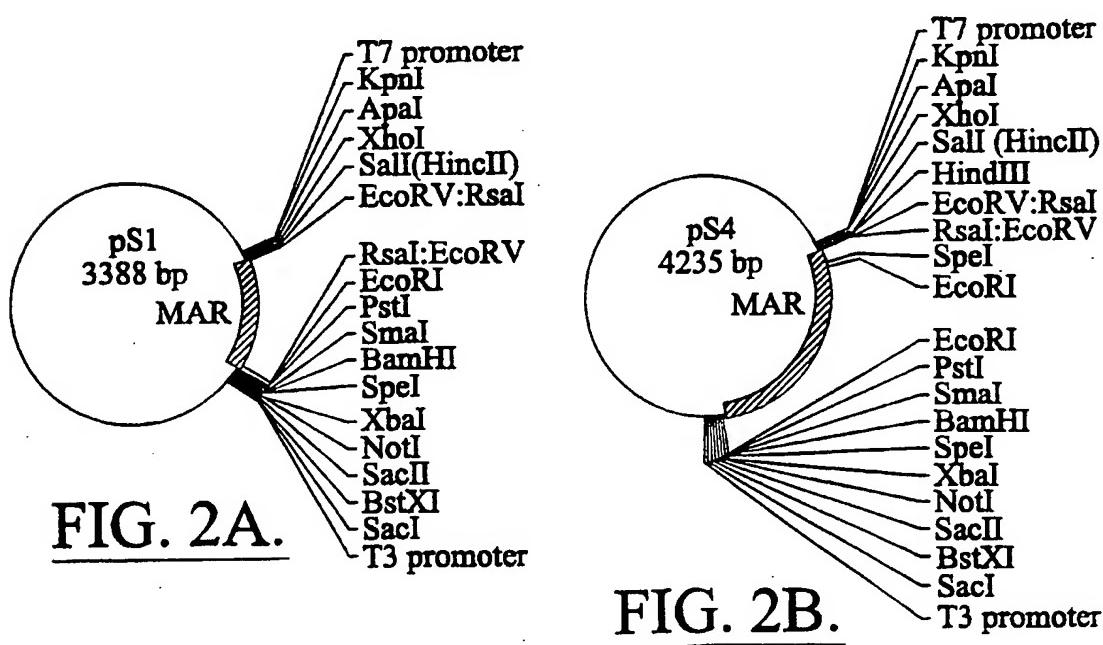
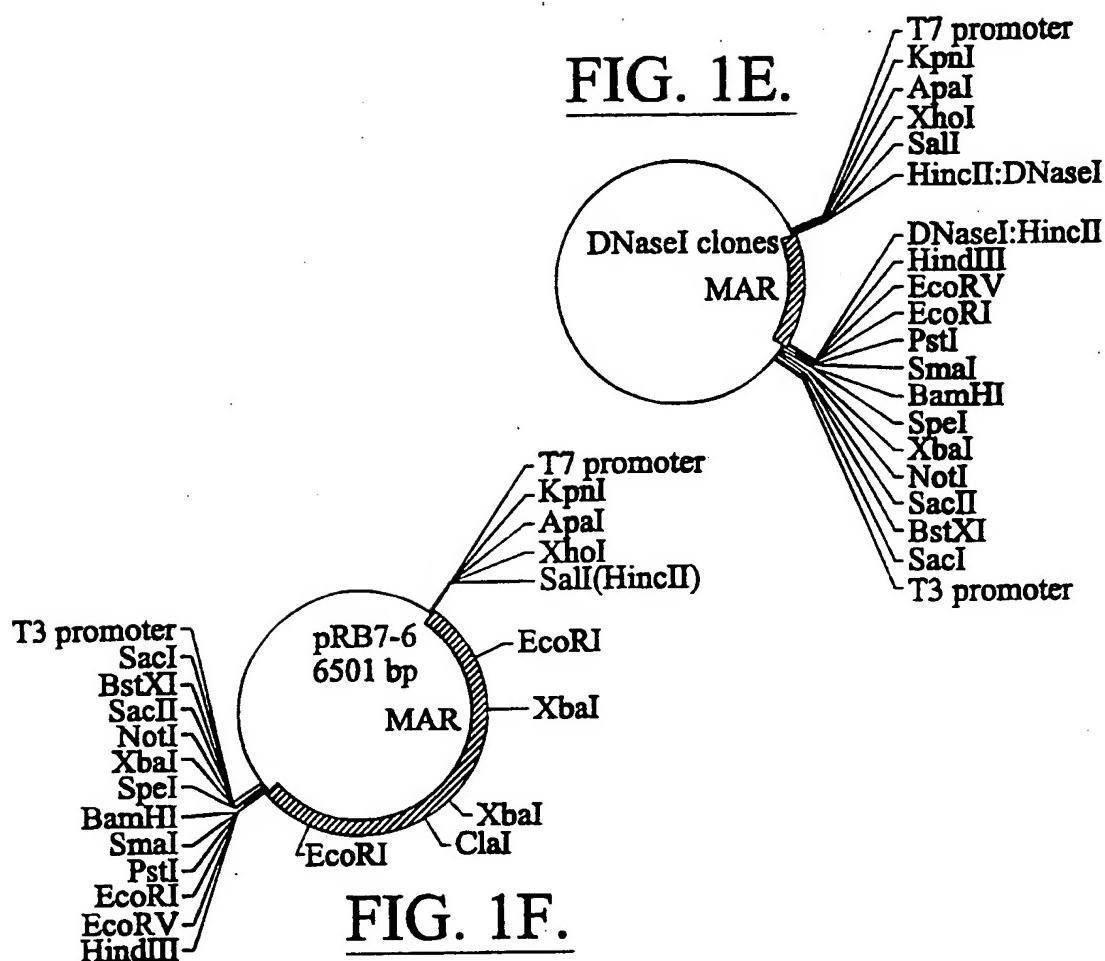
19. A recombinant tobacco plant comprising transformed tobacco plant cells, said transformed tobacco plant cells containing a heterologous DNA construct comprising, in the 5' to 3' direction, a transcription initiation region functional in plant cells, a structural gene positioned downstream from said transcription initiation region and operatively associated therewith, and a matrix attachment region positioned either 5' to said transcription initiation region or 3' to said structural gene,

wherein said matrix attachment region has a sequence selected from SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:13.

20. A method of identifying matrix attachment regions in a DNA molecule of known nucleotide sequence, comprising identifying a sequence section of at least twenty contiguous nucleotides that is at least 90% A or T nucleotides, wherein the presence of such a sequence section indicates a MAR encompassing said sequence section.

21. A method according to claim 20, further comprising preparing a MAR molecule of at least about 300 nucleotides, said MAR having a sequence which is a contiguous fragment of said DNA molecule sequence and which encompasses said 5 identified sequence section of at least twenty contiguous nucleotides that is at least 90% A or T nucleotides.

**FIG. 1A.****FIG. 1B.****FIG. 1C.****FIG. 1D.**



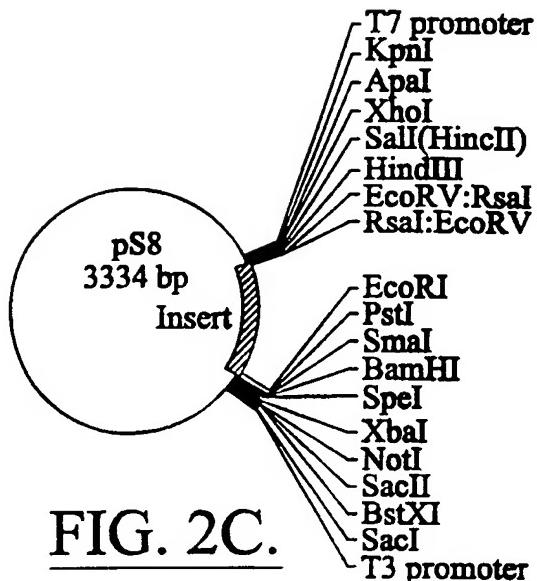
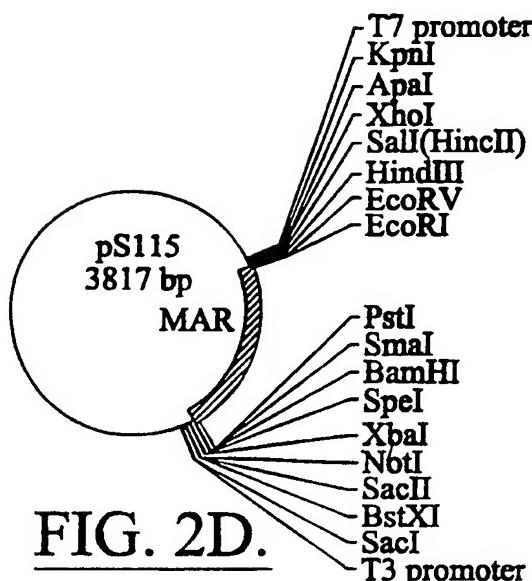
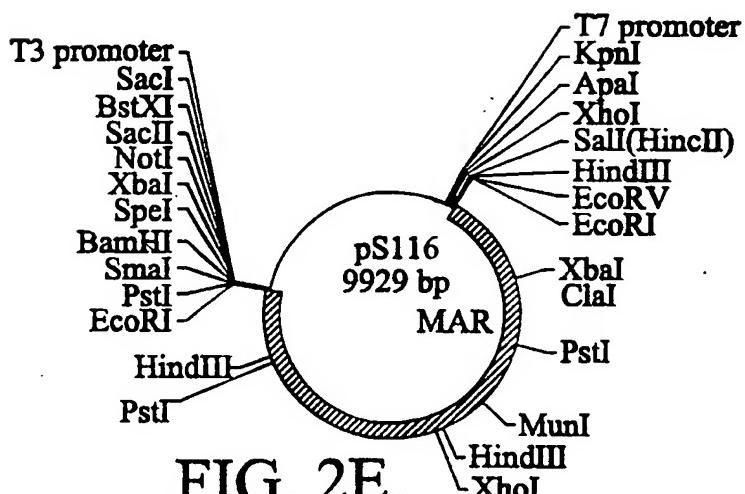
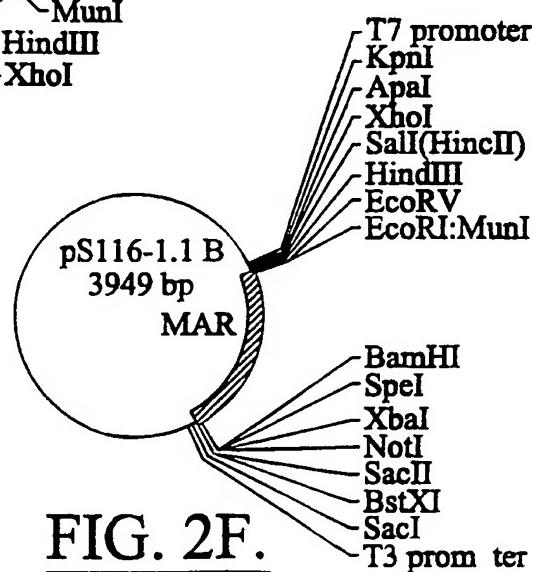
**FIG. 2C.****FIG. 2D.****FIG. 2E.****FIG. 2F.**

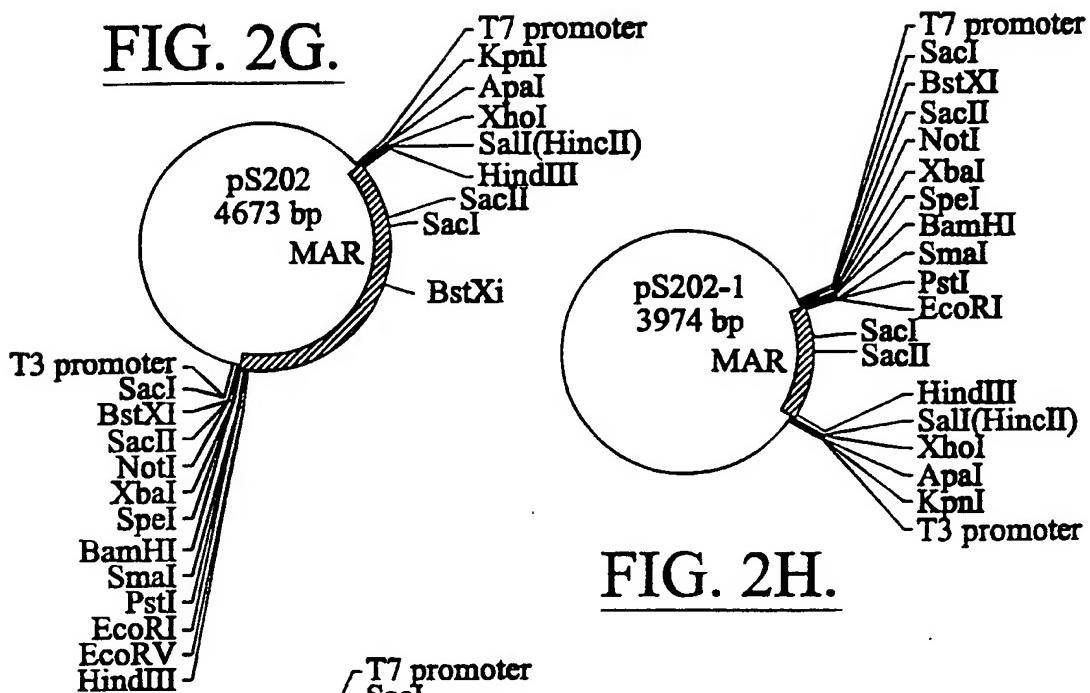
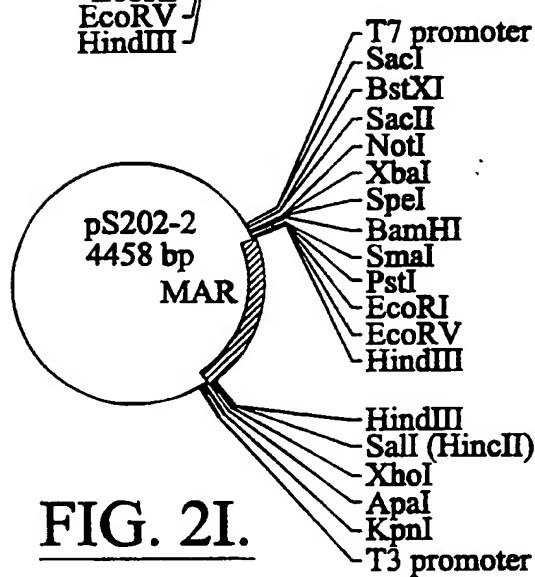
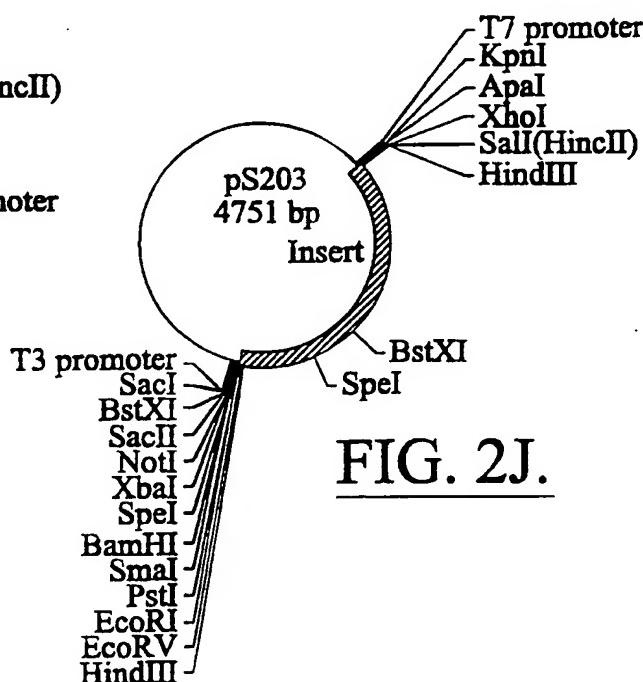
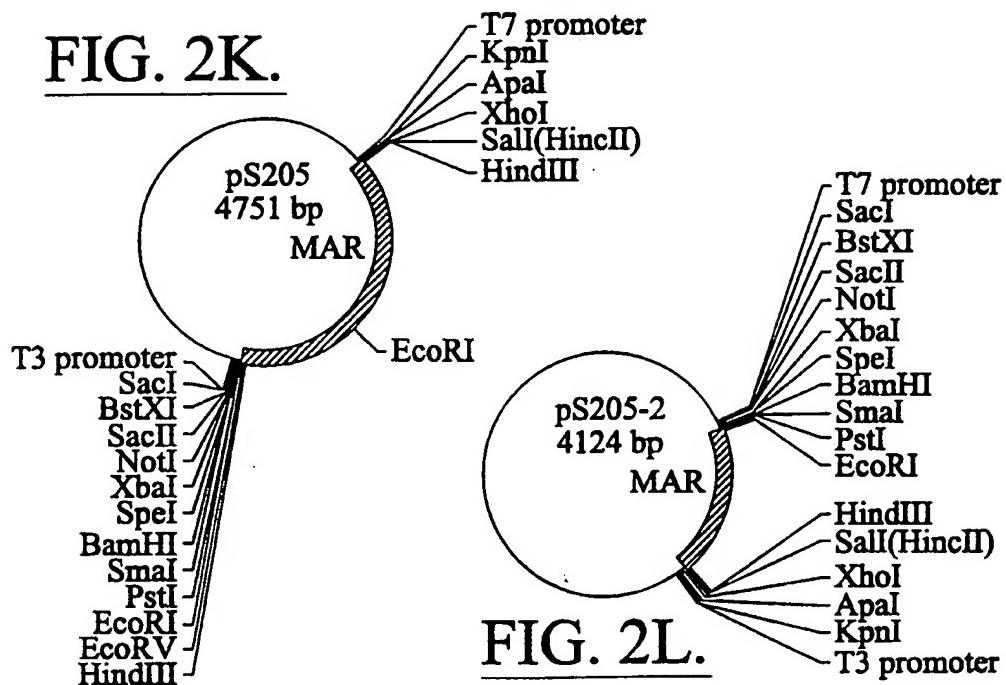
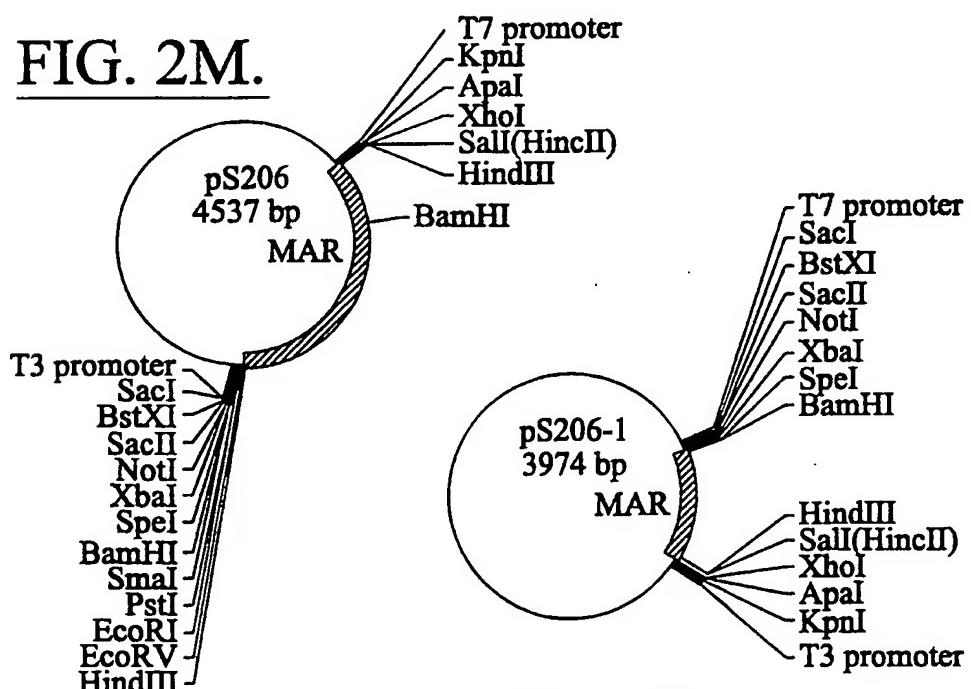
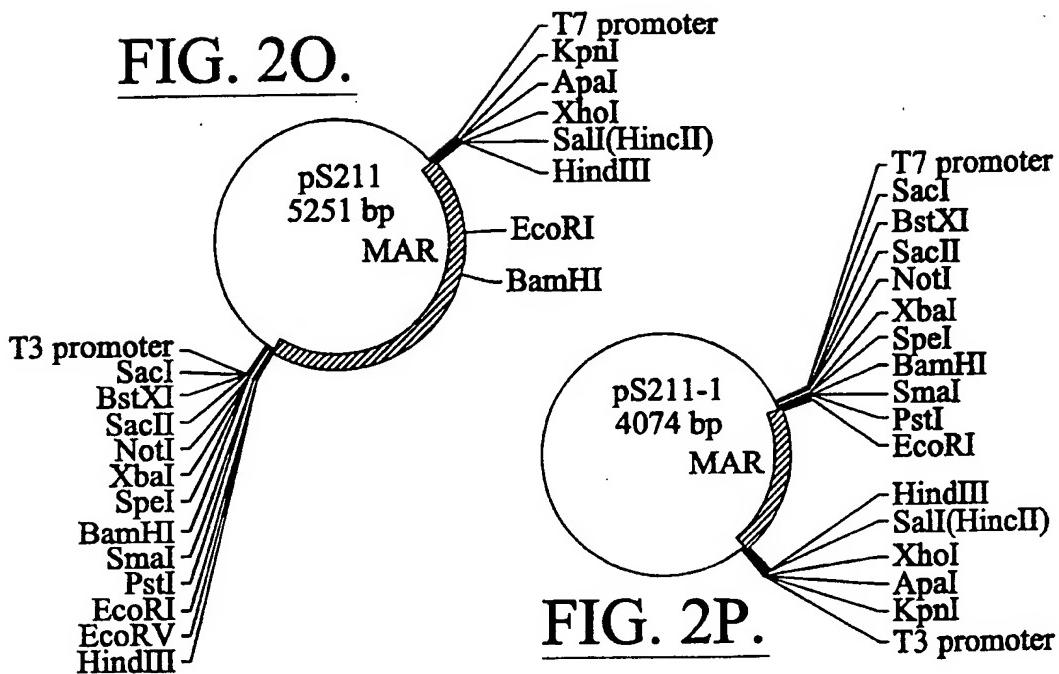
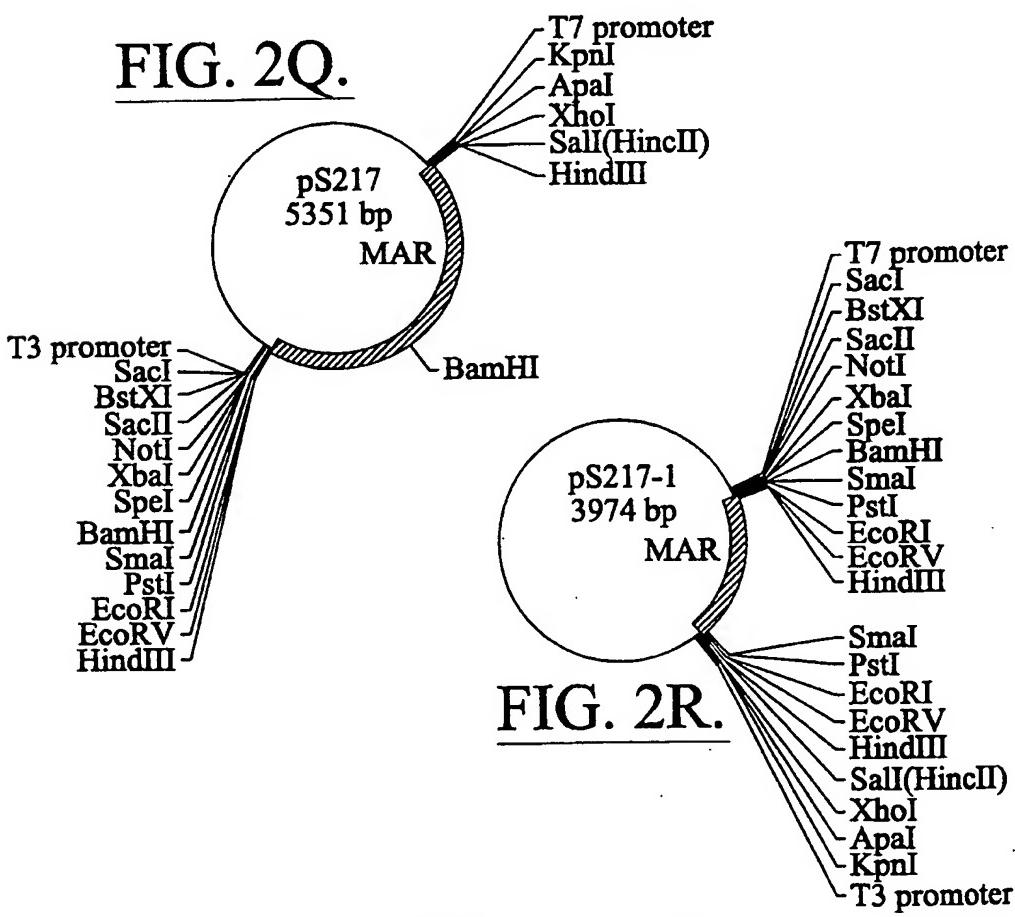
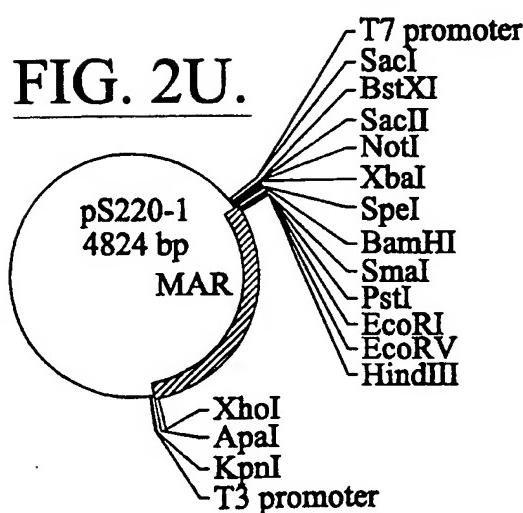
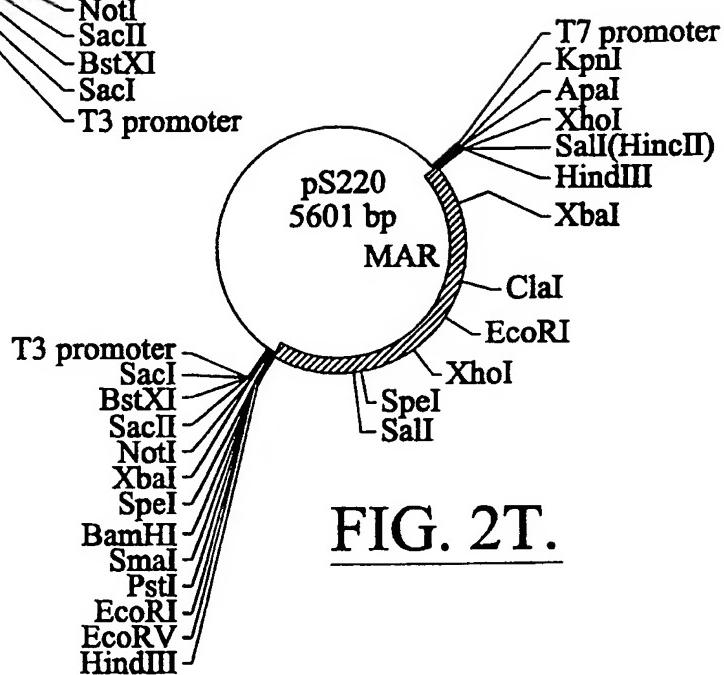
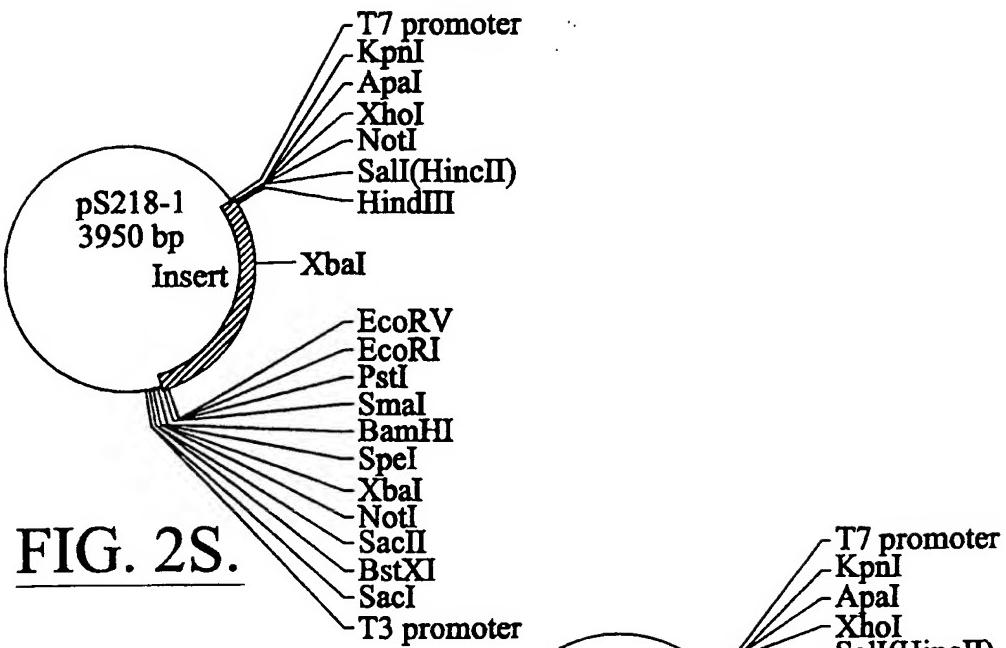
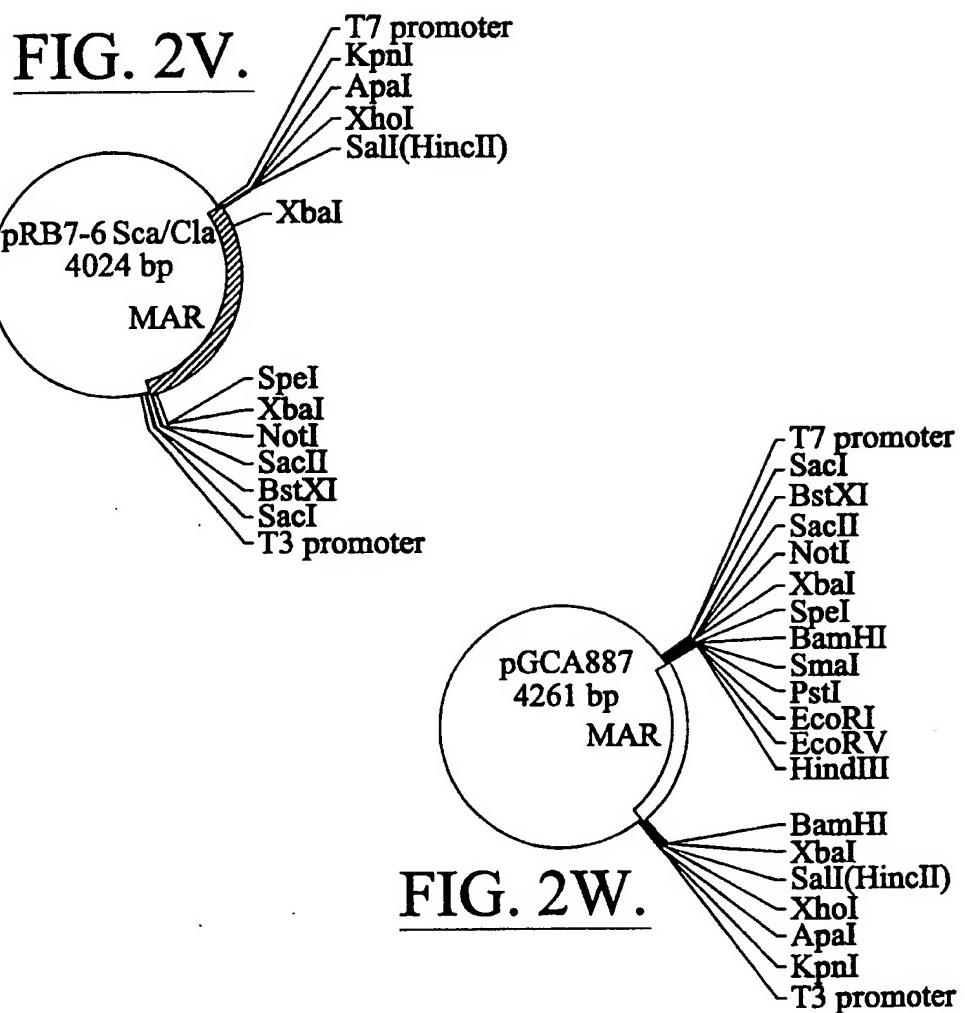
FIG. 2G.FIG. 2H.FIG. 2I.FIG. 2J.

FIG. 2K.FIG. 2L.FIG. 2M.FIG. 2N.

**FIG. 2O.****FIG. 2P.****FIG. 2Q.****FIG. 2R.**





9/18

PS1  
(SEQ ID NO: 1)

1 GATACGTAAA CRACGTGTAT CCAGTAACTA TCAAGCCTAA TCTCGAAGTG  
 51 GTAGAGACGA GATGACCGAC TTTGACACTC ACTATGGGTC AATAATAATA  
 101 ACTGAAATAA AACTAAGATA TTTAAACCAA CATGATTAC AGAATTACAA  
 151 ATAATTTATT TAATCAGCAG AAATAATCAA ATTTCTTCAA ATGTAACAAT  
 201 TCTCAATATA TTAATTAAAT TCCTTCATT CAAATAATT CTAATTTATC  
 251 AATTAAACCT CATTACAGG AGTAACAATT AATTCTTAA CAAGCAAGAA  
 301 TAATAATTCA TTAAATTCCA AGGATTTTC AATTATTAA TAGCTTCAC  
 351 AACCTGAAAT AAATTATTAA AGTATCGTGT AATTATTATT ATTAAGCACG  
 401 ATTTCTGCCG AGGACATACG GCCCGATCCA GAGTATC

PS4  
(SEQ ID NO: 2)

1 GATACTAGAG TGGTGTATC AATTCTTACT CGTATGAATT AATTAAATT  
 51 GTCTCTTATT TCTGTCCTAA GTCATATACA AGAAATGCTA ACTCCATCCG  
 101 TTTCAATCCC TATGACATAG TTTGATTTGA TTGAATTGAA AAATTAAAGA  
 151 AACAAAAGAT AATTTTGTTG ACTCATAATT TAGACATGTG TTATAAGACT  
 201 TTTCTCATGA ATTTTTAGA AACAAATGAT AATTTTGGA ACTCATAATT  
 251 TAGACGTTT ATAAAAAATA CTAAC TGCACT GTGTTCAAT ATTTATGTGT  
 301 TATTCTTATA AAACCTCTGG ACTTATATT TTAAATATT CATAATATT  
 351 GGTATCGGTA TAATTTTTT GTCAC TTTTG GATGAAAGGG AAGTTAAGT  
 401 AAATTTCTTT TTCCAAATT AGAAAGTTAT AATATTCTTT TTAAAACGCC  
 451 CAAAAAGAAA AATAAGCTAT TGATTATTAT AAGCCTAAAC CAAAAGAATT  
 501 CTTTGACTAG TAGGAAGCCA TTTTAAGTT AGGCGCCAAA ATTCAAAGGCC  
 551 AACGTGGGCA TATCTCCAAA CTGGCGGCTA CAGTATC

PS8  
(SEQ ID NO: 2)

1 ACCGCTTTA TTATTATTAT TTTTACCGAG AATTACAACA TCATGAAAAT  
 51 ACATCTCGAA CCACGTACAA TCAATGCACC CGCGGTTATT GACATATTTC  
 101 AACTCTGTTG AGATTGGAAT TTGGGTCAAA TAAATGTGCA CCCGAGTTA  
 151 AGAGGATAAC ATTATTAAT ACACGCTAA AACGACTAGC GTATCATTAT  
 201 TTTGGGTAGG GCCGTAAAT TTTGCTAAAC TGCCCATCCA GAAATCTAAG  
 251 TAATTTTACCA AACACGTATA GAGGGCCCCA CAGCTTGTGT ATTTTGTTT  
 301 GTCGAGGCTC GTCTCATTC TTATTTTAA AAGGAATTG CAACGTCTG  
 351 GAAATGCATC TCGAACCAACG TCACAAATCAA TGA

Fig. 3

10/18

PSI15  
(Seq 10 No: 4)

1 GAATTCGATA GACTCACTTA AATATTAGAA GTGATTACC TAGAGTTAGA  
 51 TCCAAAACAA TTATCTTGCA CCTATCCTAT CAACCCCTAT CTTTCCCAT  
 101 TGATTACTAC CTTGCTTACC TTTGTTACGA TTTCATTAG ACAATAACTT  
 151 TAGATTCTTA GTTAATTGCA GTTAGAAATT ATATTAAATT TCAATTGTTG  
 201 GATCATCTTG AATACCAATC AAGCTAGAAA ATACAAGAAT ACTGTTAAA  
 251 TCAAATCCAT GTGGATACGA TATTATACTA TATTATATTG GACTTGTGAG  
 301 CATTATTAT GTGTGTTTG TGCTCGTCAA AGTTGGCGT CGTTGCCGAG  
 351 GATTGGCAAT CAATAGTGT TGAAATAGTT TTTGGTGCTA ATTTAGGAAT  
 401 TAGGTTTAT TTATTTATT TTTCTTTCT TTTCTTTTC CTTTCTATT  
 451 TTATTCCTT TATTAGTTAA CTTCTTTCA AGATTTTTT TGTAGTACCT  
 501 ACAAGTTAG AGAAGATACT GTAGATTTG AACTCTAAAT GTTGTGAAGA  
 551 TGGAGTACAA CCAGCCTAAG AAAATATTG AATAGTTAGC AGCTGAACAT  
 601 TATCGGCGGT CGGTTATGCG GTTAAATGC GGTGGAAGCA TCTACCACCG  
 651 CAGCCTAAAG AAAATATTG GAATAGTTAG CAGCTTGAAC ATTATCGGCG  
 701 GTCGGTTATG TGTTTAAAT GCGGTGGAAA TCATCTACGG GCTAATGTC  
 751 AAGCAGGTAT GTATTCTTCC TATGGTTCGT ATTTGAGGA GTCTCACTCT  
 801 GTTTCTAGTT CGTACATGTA TGAGGATTCA TATGGGCACA ACTCTGACTC  
 851 TGGTTGGGAT GAATC

PSI16-1.1B  
(Seq 10 No: 5)

1 GAATTGTATT ATTGTTAGGT GGGAGAGATT TTTGACTATA TGGGTTAAAA  
 51 TCAGCGACAA AGGGCCAAAT ATACCTATT ACTTTAAAAA ATAGTCTAAT  
 101 AATACCTCTC GTTATATTAT TAGGTTATCT ATACCTTTGC AGTCATATT  
 151 TGGGTTCAAA TATACCCCTC ATTAAACGG AGGGACACGT GTCATCGTCC  
 201 TGTTGGTCAA TTCTAAATAT CTCCCTAATTA ATTAAAAGA CTCATTACCC  
 251 ATATCCGAAA AATATTTTT AAAGCAATAT TTTTTATAA AAAATGGAAA  
 301 AACTGAAATT ATTTTACTA AAAATTGAAA AAAAGAAAA TAGTTTTTT  
 351 TCAGTTTTA CAAAAAAACT ATTTAGAAA AAATTGAAA ATATTTCTA  
 401 AAACAATGTT TTTGTAAAAA CTGAAAAAAA AGAAGCTGAA AATCAATT  
 451 CTAAAGCAAT TTTATTTGTA AAATCTGGAA AAAACTACTA AAAACTGAAA  
 501 AAATGAAAAT ATTTTTTTT CTAATTTTA CAAAAAAAC TGCTTTAAAA  
 551 AAAGCTGAAA ATATTTCTA AAACAATATT TTTGTAAAAA CTAAAAAAAAA  
 601 AATATTTCT TCTTTTTTC AGTTTTTAGT TAAAAATATT TAAGTTTTT  
 651 CCAGTTTTA ATTACTTTAG AAAATTACTT TTCTGCTTT TTTTCAGTTT  
 701 TTACAAAAAT ATTATTTAG AAAATATT TCAGTTCTTT AAAGCAGTTT  
 751 TTTTTGTAA AAACCTGAAA ACAATATT TCCTTTTTT CAGTTTTAG  
 801 TAAAATTGT TTTAGTTT TTCAGTTT TACCAAAAT AAAATTGCTT  
 851 TAGAAAATTA TTTTCGGGT ATGGGTAAATG GGTCTTTTA ATTAATTAGG  
 901 AGATATTTG AATTGATCAA TAGGACGATG ACACATGTCC CTCCGTTAA  
 951 ATGAGGTGTA TATTGAAATC CAAAGTAAGA CTGCAGCCCG GGGGATCC

Fig. 3  
continued

11/18

PS202-1  
(Seq ID No: 6)

1 GAATTGATA TGGCTTGTG GACARGAATT AATGAATCAA TTGTGAAAAA  
 51 GTTGATGGAC ATATTGAAGG TAAAATCATA TACTATTTT CTAAAATCTC  
 101 TTTTAAATGT TCCCCAATTA TCTGATTCT ATATTGCTCT TAAATGTCAC  
 151 TCAACCTTAG ATCAACAAAC ATATAACTTA CCCAGTACAT AAGAGATTGC  
 201 GGCATTATGG CTTGAAGAAA ATCCTAGAGA CACATCTGCA CCACATATT  
 251 GAATTTATAC CCACAGTAAT AGAGCTCGGT TAGTACATTA TTATTATGGA  
 301 TGTTACGATC CGTTGCAGTA TCCATTATTA TTTCCCTCG GTGAAAATGG  
 351 ATGACATTGT GGAATTAAAA AAATTATTCA GACAAAAAA TCGACGAAAC  
 401 GTAGAGCTTA CTGCGAACAT GAACAATTGC CCAGTATATC AAATACGTGT  
 451 TCAGTTGATG GATTCTTGA TATGGAAGAT GAATCACTAC AAAGAGGAAA  
 501 ACGAAAAAAGA GATACTGTT CTTGTCGAGA GTATTATTGT TACAAATTTC  
 551 AACTAAGAAA TAATGAAACA AATGAAGTGT TACATTGTGG GAGAATATTG  
 601 CAACAATTAA TAGTAGATAT ATATATATAA AGCTT

PS202-2  
(Seq ID No: 7)

1 AAGCTTGCAC GCCTACATCG TGGGATAATT TAGAAAAAGG AAAGGGTATA  
 51 TTGGATCCCC CTATCATTTG TGAAACAGGT AACCATACGA GAACCCCTT  
 101 CGCTTCCTGA AAAATGTTAT ATATTGTTGT ACTCATATT ATACACTATT  
 151 TATTATTAAT ATAACGATGC TTATTTGCT TGGAGATTGG AGATTATCAC  
 201 AGCTTATTAA TCTTATATTG TATCTTATTA AACCTAAAAA CATAAATAC  
 251 ACGTGCTCTT TTAATTGGG ATCTATTAAAG GGTCGTTGC ACGCTTTAA  
 301 ACATCTTGGC TATTCTGTT ACCAGCTGCT ACCTTAGCCT GTATGCTTAC  
 351 ATCATCTCCT AATTAGACA AAGGAAAGGG TATATTGGAC CCCCCCTATC  
 401 ATTCTGAAA CAGGTAACAC TACATTAGA TTATACCTT TTCAGAATGA  
 451 CATATTGTT ATACATTACT GTAAATTGTG ACTATTGTG TATTAGGGTC  
 501 CACATCGGGT ACATCTAACCG TGCGTCATGT TATCTTGAAC ACTGTTCAA  
 551 TCAAAGGTTT GCACAAACTT AATGTTACAA TCATGTCCAC CATACTATG  
 601 CCTTGGTGCT CTTTTTTTC CTAATGATAC TTCTTATATA TTCAGCTCAT  
 651 AGGCAGGCCA GAAAGGTGTG CCTGGTCACT AAAGAGCAAC GAAGTGAGTA  
 701 TGTTGCTCTA AAAAGGTCC CACACTGTCA ATTCTGTCA CCAAAGAAGT  
 751 TTGAATATGA ACCTCCAGGA TTTTGCTGTA ACAGTGGTTC AATAAGGTTG  
 801 ACATCTCATA AAATGCCAAC TGAATTATCG GAGTTATACT TTGGAAATAC  
 851 TGAAGAATCT GAAAATTTC GAACTTATAT TAGAACATAC AATAACATGT  
 901 TTGCATTAC TTCACTGGT GTCAAGTATG ATAAAGAGCT AGCGAGAAGA  
 951 AATTGTGGTA TCTACACATT TAGAGTCCAG GGACAGATGT ATCATTAT  
 1001 AGATGATTTA GTTCCTCCA ATGAAAAACC TAGGAATTAA TAGCTGTACT  
 1051 TCTACGATAA TGATAATGAA CTAGCCAAAT CAAGCTT

Fig. 3  
Continued

12/18

pS205-2  
(Seq ID No: 8)

1 GAATTCTCA GCCATTGTAC ATATAGTTGT GTATTAATGT TATTAATAAT  
 51 GGATAATTAA ATATATACCT GGAATAAATA TACGATATTA TAATAGTGTG  
 101 TAATTATATA TAAAAATTAT ACATAATATA ATGATGGTAT TTAAATATAGC  
 151 ATAATTGGA ACGATCTGGA TTGATTTCTT GAATCAAAT AGAGTTGTGT  
 201 GAAAAGAAAA GAATGAGATG AAAAGCAAAG TATGAAGAGA TGAATTGTTG  
 251 TTTTTTAT GGAGGAGGAA GGTTCTCAGT GATGGAATCA TCCCTGGTTT  
 301 TCTTAGCAC CAATGAAAGT AATGAACCCC CCCCAAAAAA AAAAAAAA  
 351 AAAAAAAAGG GAGAGAGAGT AGAATGGAAC GGCTAGGTGA AAGTATAGGA  
 401 GTAGAAATTA GGTCAGGGAA GAGAAAAGGG GGGAAATTAA TTCCTAAATT  
 451 AATGGGATTC TAATTTTAA ACTGTTTGAA AATATTTAA AAGTAGTGT  
 501 ATTTATATTA TTAACCTTTA AAAAAAGTCA AACGAGGTAA AAATTCCATG  
 551 GGGGAAATT TAAATGGTTA GTCTCTATA ATATTTCAA CTCTGCTTAG  
 601 CACTAAAAAT TAGTCTAAAA ATAACCCCTAA ATTAGTGTAT CTAATTAAAT  
 651 TAGTTCATCG AACAGGAGCA TTGGATTATC CCTCCAGAGT TACACAGGAA  
 701 GCTT

pS206-1  
(Seq ID No: 9)

1 GGATCCAGCT ATTATTATAG CATGTGAGTT GTCCGTGAAC AGCTAATT  
 51 TTACCAACC CAAATTCAAT ACTATTTAG TGTAAATATA TCTTTAGGT  
 101 CTAGTCTTAA TATTTAACCTT TTTGTCTTAC TTTAATAGA TTTTATTGAA  
 151 GAAAATTAA TAATTACAAA AAATAAAAAG TATATATTCA CATACTTATA  
 201 GTACAAACTT TGTTCTATT TATAAAGAGA AAAAGAAATT TTACAAAAA  
 251 CAAATATATT TGCTTCTTT TAATTAGTAG TTTTATTAAAG CAAGCTATAG  
 301 AAGCTC

pS211-1  
(Seq ID No: 10)

1 GAATTCCGTG GTTTAGCAC GGTGCTCAA TTGTCATATT TGGCTCATTT  
 51 ATCTGATTT TAAACAATTA AGAACTTATA TGCAAATTAA ACTTTAAAA  
 101 CCGCTTTAT CATTATTAT TTTATACAAA ATTACAACGT CGTAAAAGG  
 151 CATCTCGAAC CACGCCACAA CCAGTGCACA CGTGATTGT TGACGCATT  
 201 TGGACTTCGT CAAGATCGTG ATTTGGGTTA CATAATGTA CACCCGTAT  
 251 TTAAGAAAAT AACCTTATTA AATATTGCGC CAAAATACTA CGCGTTATGA  
 301 TACTATTAGG GTAGGCTTGT GAATTTTACT AAATGCCCA TCTCGGAATC  
 351 TAGGTATTT CTTATATTAA AAAAATAAG ATGGGGGCCT GCAATT  
 401 ATTATTTAAT ATTTATTTAT TTTTAGCGA AGATCCCTCC CTTATTTAT  
 451 GAATACCCCTT TAATGACTAC ATCTTATTA TTACTAAGTT TGTCTATAAT  
 501 TATGAAGTCA ATCTCTACAT ACATAAAAAT AACATATTAA TTACTAATT  
 551 AAAACAAATA TTAATGGAAA GTAATATTAC TAAAATTATA ATTACAAACA  
 601 ACATGGAATT GTCACAAAAT AAAAATAAA AACTAATTAT CCCATAGTTG  
 651 GATTAAAATT CATATTGTTA GTATGACTTA AGCTT

Fig. 3

continued

PS217-1  
(SEQ ID NO:11)

1	AAGCTTTAAA	AGGAAGAGAG	CCACAA	TTT	CTTGACCTT	CCTTCTCTCC
51	TAGCCACTAA	GATATA	CACT	ACTGGTC	AAGAGCATAT	TTATAGCTCA
101	AAATTG	TTT	CTGT	GTAAACGT	<u>GA</u> TTGTTCTTA	<u>CTTGGATTCT</u>
151	TGTTCTATAT	ATTTACGGGA	GAAAAGAGCA	ATTTGCATGC	TCCTAAATCT	
201	TTTATTTCT	GGT	AAAAAAT	TGGTCTTAA	TTGGCTGGGA	ATTATTTTT
251	AGATGCTACA	ACCTTGACAA	ACACCTAAGA	ATATTTAGT	GACAATGGCT	
301	TGTTCTTGA	GTACTGGTT	TTCTGTTCT	GGTCCTGTT	TCAACGCCAC	
351	AGCCAAAGAG	TCTCGTCGTC	ATTGCCCTTC	GATTGGCACT	CTGCAACTTA	
401	AAGATTTAGC	ATCCAGAGAA	TTTCTAGGCA	AACCCTGGA	TTATGCATCA	
451	GATCATATTG	GTACTAACCA	TTGGAATGTT	GAACGACTTT	CTGTATGTAA	
501	ATCTCTGATA	CATTGCTTC	TGTGTTATA	CTTGGTGT	TCATGTTTC	
551	ATTCTTGT	TAATTTTC	GAGATCAAAT	CATTATAAG	TATTTATTCT	
601	AATGATT	GGCACAAAGTA	TCAATCGCTG	CTCAGAGATG	GTGGGAGAAG	
651	ACCCTTAAAC	CCAACATGGT	AGAGATCAAT	TCAGCAACAC	AACTTGTGA	
701	TTCATTATTA	<u>AAAGCTGGTA</u>	<u>ATAGATTGGT</u>	CATAATTGAC	TTCTTCTCTC	
751	CTAGCTGTAG	AGGTTCAAG	ACTTACATC	CTAAGGTAAG	ATATATAGCA	
801	ATCCCCTAAA	AAAAAAA	AAAAAAA	AAACCAACAA	CTACATCGTA	
851	ATCCTAAGCA	AGTTAGGGTT	AACTATATGA	ATCATCACTA	GACGGATCC	

PS218  
(SEQ ID NO:12)

1	AAGCTTAAC	TTACTCACAT	TGCTTCTTT	AGGGAAAGCGT	CTTCTTAAAT		
51	GACCAC	CCTC	AAA	TTCTC	ATGAATCTC	TTCTGTTGTC	CACTCTGTTA
101	TCGCTGAAAC	GAAATCTGAA	ATTGT	CATGA	TGCTGACTAT	TATCCAATCA	
151	CTCAGTCTCT	AATT	CATATT	TAGATTATCT	<u>TGTT</u> <u>CA</u> <u>CCAG</u>	<u>CCC</u> <u>AT</u> <u>ACTGA</u>	
201	TTTTATTGT	TTTGGGGTCT	AACT	TTTCCT	TCCGGTAGTC	GGTTGGAGTC	
251	ATGAAC	TTT	CTGAAAT	GAGGATATGA	CTT	TATGGCC	TATACTCTT
301	TGGTGTCTCA	AGGC	CTGTCA	CCTCTCATCT	TTT	CCTCAA	TTGACTATAG
351	ACTCTGTAAT	ACT	GTCATCT	TTGGGATCTA	CCG	TTGT	CCATGTATCA
401	TATCTTA	ACTC	ATAATGCTTC	ATTA	ACTATT	TTCTTATTTC	CCGCTAACAT
451	TTATGTCTAT	CACT	TTATTC	TGAAA	ACTCG	AACAAGACAT	TCTTTCGTT
501	TTAGATCCCC	TTT	GCTCCAT	CCAGTGGT	TC	CGGGGGAC	TTAACGTTCT
551	CGCTCTCCTA	GGG	GAGGCGAG	CCACACTAAG	GTA	ATATT	TCCCTTCTAG
601	GCTTCCGTG	CCT	ATCTCT	GAGA	TATT	TTTCATGCTA	ATATTCA
651	CTAATTGTA	TTT	CTAGAG	TGCGCC	CAT	GGGTGC	CTCAAGAAGAGC
701	TATTAGCATC	TTT	GTAATAT	CCTCGGAA	TGT	CAACTAA	CACAAACACAA
751	TCCATT	CA	TTTG	GGTT	ACTCTAAC	CAGTCGGATA	CTAATATCCT
801	GTCATT	TTA	AAACTACAC	ATGTTAG	<u>CCC</u>	<u>CC</u> AATAGGAT	ATAACTAAGA
851	TGGGTGTGGC	CAATT	C	TACA	TACATCTG	ACTGTTGAA	GTAAGTCGCA
901	ATGCTT	TTT	CTGCCG	GAGTTGAA	AA	TACCGATA	CTATATTAAC
951	TGGGTAC	CTC	GTAC	CC	TTCT	ACTTG	TTGAAGCTT

Fig. 3

continued

14/18

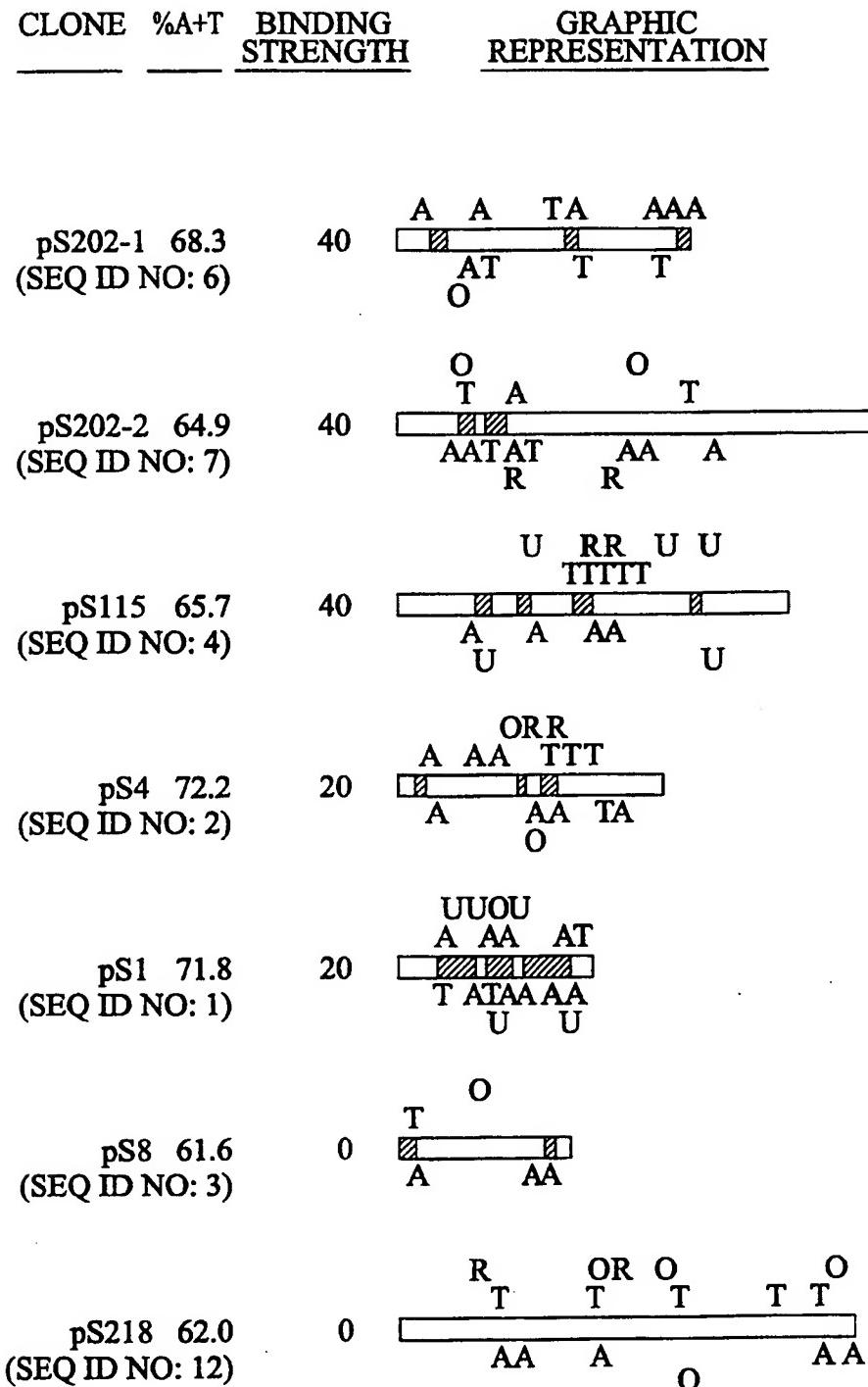
PS220-1  
(seq 10 no. 13)

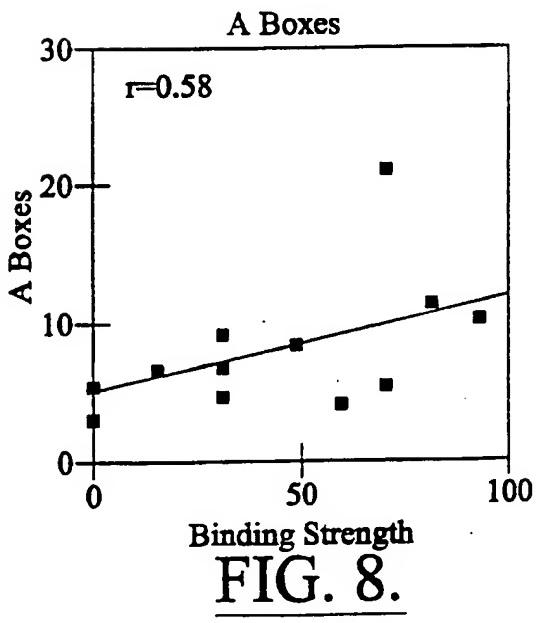
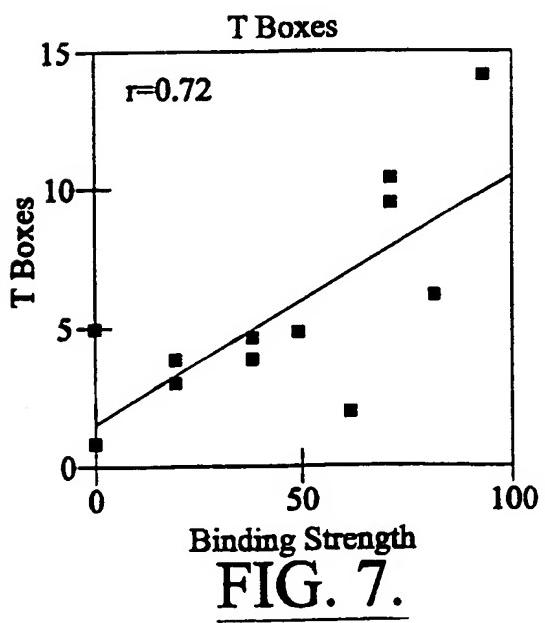
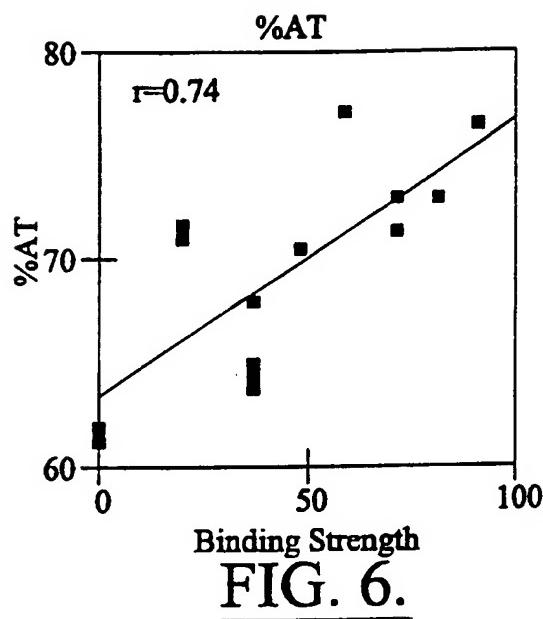
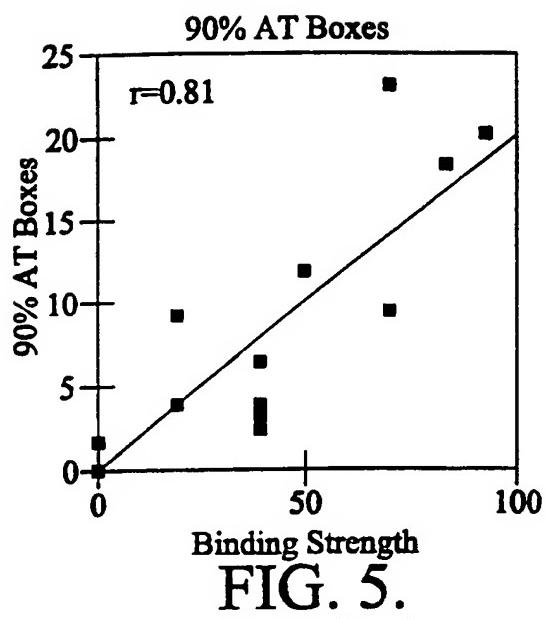
1	AAGCTTGAAA	AAGAAGAATT	AAGGCTTGCT	TTCTTAATTT	TTAAAAATA
51	AAAATTATTT	TGAACATATCT	ATACTATATT	AAAAGCACGA	<u>AAACCCTATC</u>
101	<u>GAAATGTCGT</u>	<u>TCGCCTTTT</u>	TACCCCTTTAA	AAATAATT	ACATTAGACA
151	AAATAGTCAT	TTTACTATTT	TCCTAATAT	ATAGGATT	AAAATTAAATT
201	TAACTTGGC	TATTAACAT	TTTCTTATAA	CTTGAAATAT	GTAAAACCTCC
251	TAATATTTAG	AAATTTAATT	AACATAACCA	AGGATTTTA	TATCGGTAAT
301	AACTCTAATA	TGGTATCCAA	ATCAGTCTAG	<u>AACTCTCTTA</u>	<u>CCTCTAATAA</u>
351	GTAAAAGTAC	TTCTAATAAA	TTCATATACT	TTTCTCTCT	TCTCCGATCT
401	CTCTTGCTC	TTCTTTTAT	GTATCCTTC	CTTCTAATA	GCCTTTATG
451	AGAAGTAAAC	TTTAGGGTT	GGCCCCCCT	CCCCCCACAA	TTATATAGTT
501	TCTTACTCAG	TTGTTGGAAT	ATAATTCAA	TTCTTAAATA	ATTGACGGTG
551	ACATTGAGTT	TTACTTGTG	GAAGAGAATT	AGATTCTCGT	GTTAGTAAA
601	TCGGTAGTA	ATTGATGATG	CATTATTTT	ACTCTATAAT	AGAGATGCAA
651	TTTATTTTT	GCATTTGGG	ATCAAATTGT	AATGCAGTC	TATATTGATT
701	TCATAAATGT	TTGGGATATT	GTTGGTTATT	TAACTAGAAA	TAGACTTCTT
751	ATTCATATT	TATTGTTAAA	ATCCTTTATT	GGAGATGAAT	TATTGTTCA
801	CCGATTAGAA	GTTGATAGTC	GCTTTGTT	TAGAAGAAAT	TTTACCGTAG
851	ACCAAGTTAA	GGAGTTTAG	AAGCACTTG	CATGGGAGCA	TTAGTGTATG
901	TTATGGCTT	ATCAAATATA	GGTTTGAAG	ATTCAAGAG	<u>CCAAGAAAAG</u>
951	CTAGAACCA	AGAACTAGGA	AGTTAGAGTA	<u>ATTCAACAATA</u>	<u>CCATAACGTG</u>
1001	ATATAAAACT	TTTTATTGTA	ACTCAAATCG	GTAATATT	TTGCTTAGT
1051	CTTAATCGAT	AAATTATTT	TTTATATTGA	TTAGTTATAG	GAGGCTCACA
1101	AAGTTGGAA	TAATTAAAAT	ATCATATT	GTATTTGAAC	AATTATGAA
1151	ATAGTAATTG	GTAAAAAATC	ACTTTAAATT	TTTATCCTAT	ATCCAGAAGG
1201	ATTATGGGT	CTGGCATAGT	TGTTTGGAG	ATTGAATCA	GGGTAAAAGT
1251	ATGTTGTAAT	TTTTATTTG	TTATAGGCAT	TTTTGTGCT	TGATTGTTT
1301	GTTGTCATTA	TATTTATTA	TTTGGAAAGTG	TATATATATG	TTGATTAAA
1351	ATATAGATAA	TCAATTAT	AAGAAATTG	CAACAATTAC	ACAAGGATAA
1401	AGTCTACAAT	ATGCGAGTAA	AATTGATTG	AACCTAGGAT	GTCATATT
1451	ATGCATATT	TATTTCAATG	TGTTTATT	ACATCTATTG	TATTATATG

Fig. 3  
Continued

<u>CLONE</u>	<u>%AT</u>	<u>BINDING STRENGTH</u>	<u>GRAPHIC REPRESENTATION</u>
pS116-1 (SEQ ID NO: 5)	76.6	90	<p style="text-align: center;">UU U UUUUROU</p> <pre> T T ATA TAT AT T TA A T A T T A T A A T A U O U U U U U </pre>
RB7-6 (SEQ ID NO: 20)	73.2	80	<p style="text-align: center;">UORR R U UU</p> <pre> AT AA AA A T A T U T A T AAA O A T T A U T T A U </pre>
pS211-1 (SEQ ID NO: 10)	71.7	70	<p style="text-align: center;">T OR O</p> <pre> A T TAA ATT T A O O R </pre>
pS220-1 (SEQ ID NO: 13)	73.1	70	<p style="text-align: center;">U R URO RU</p> <pre> A A T A A A A U T R A T A A A A A T A A A R </pre>
pS226-1 (SEQ ID NO: 9)	77.1	60	<p style="text-align: center;">U O U</p> <pre> AA A U TATA U R U </pre>
pS205-2 (SEQ ID NO: 8)	71.4	50	<p style="text-align: center;">R R</p> <pre> AAA T AA A T A T A A T T A O O U R U R </pre>
Ps217-1 (SEQ ID NO: 11)	65.2	40	<p style="text-align: center;">OU TT RR T AAA</p> <pre> OU TT AAAA AA A TT </pre>

**FIG. 4.**

**FIG. 4A.**



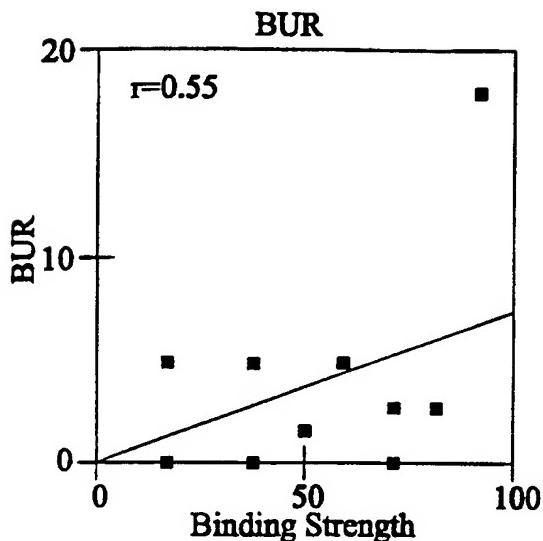


FIG. 9.

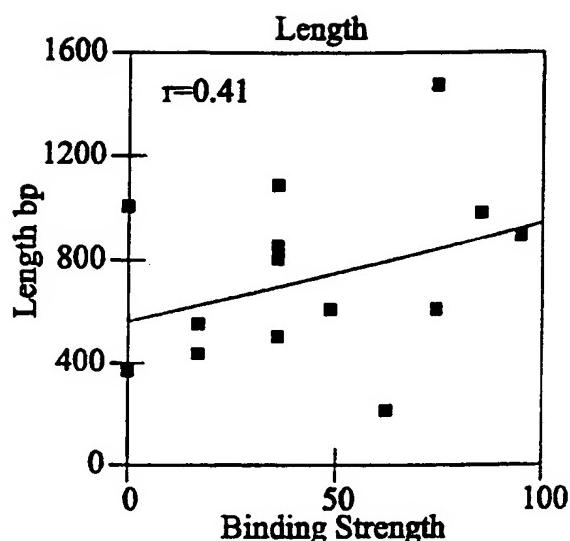


FIG. 10.

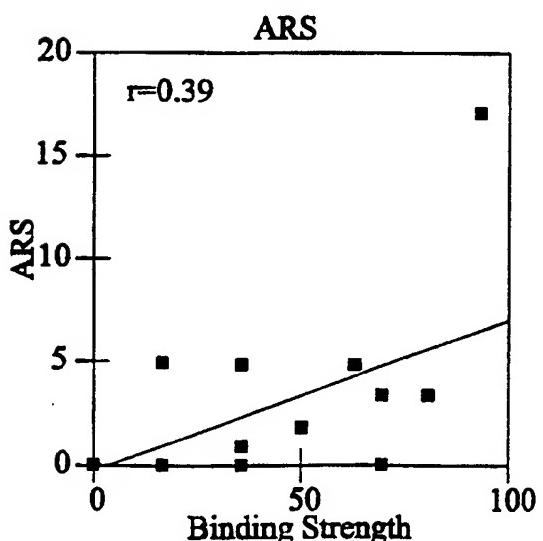


FIG. 11.

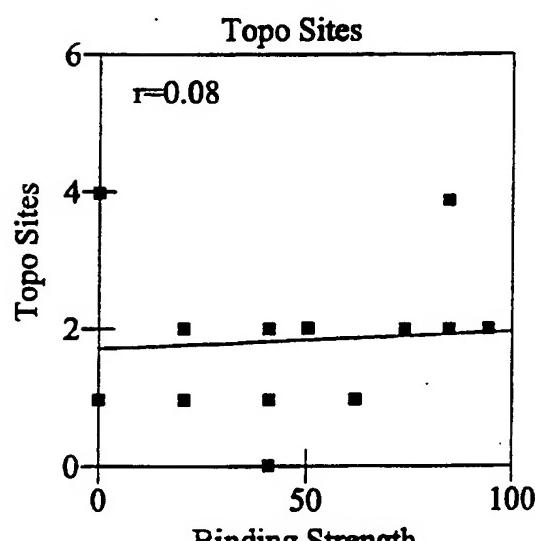


FIG. 12.

## SEQUENCE LISTING

<170> PatentIn Ver. 2.0

5       <210> 1  
      <211> 437  
      <212> DNA  
      <213> Nicotiana tabacum

10      <400> 1  
      gatacgtaaa caacgtgtat ccagtaagta tcaaggctaa tctcgaagt gtagagacga 60  
      gatgaccgac tttgacactc actatggtc ataataata actgaaataa aactaagata 120  
      tttaaaccaa catgatttac agaatttaca ataatttatt taatcagcag aaataatcaa 180  
      atttcttcaa atgtaacaat tctcaatata ttaattaaat tccttcaatt caaataatcc 240  
      ctaatttata aattaaacct catttacagg agtaacaatt aattccttaa caagcaagaa 300  
15      taataattca ttaattcca aggattttc aatttattaa ttagcttcac aacctgaaat 360  
      aaattattaa agtacgtgt aatttatttt attaaggcacg atttctgccg aggacatacg 420  
      gcccgcattca gagtatc 437

20      <210> 2  
      <211> 587  
      <212> DNA  
      <213> Nicotiana tabacum

25      <400> 2  
      gatactagag tggtgttatac aattcttact cgtatgaatt aattaaattt gtctcttatt 60  
      tctgtcctaa gtcataataca agaaaatgcta actccatccg tttcaatccc tatgacatag 120  
      tttgatttga ttgaatttga aaatttaaga aacaaaagat aatttttgat actcataatt 180  
      tagacatgtg ttataagact ttctcatga attttttaga aacaaatgat aatttttgaa 240  
      actcataatt tagacgtttt ataaaaata ctaactgcat ctgggtcaat atttatgtgt 300  
30      tatttcataa aaactctgg acttataattt ttaaatattt cataatattt ggtatcgta 360  
      taattttttt gtcactttt gatgaaaggaa agtttaagt aaatttcattt ttccaaattt 420  
      agaaaaggat aatattcttt ttaaaacgcc caaaaagaaa aataagctat tgattattat 480  
      aaggctaaac caaaaagaatt ctgtacttag taggaagcca ttttaagt aggcccaaa 540  
      attcaaaagcc aacgtggca tatctccaaa ctggcgctca cagtatc 587

35      <210> 3  
      <211> 383  
      <212> DNA  
      <213> Nicotiana tabacum

40      <400> 3  
      accgctttta ttatttattt ttttaccgag aattacaaca tcatgaaaat acatctgaa 60  
      ccacgtcaca tcaatgcacc cgccgttatt gacatatttc aactctgtt agatttgat 120  
      ttgggtcaca taaatgtca cccgagtttta agaggataac attattaaat acgcgcctaa 180  
45      aacgacttagc gtatcattat tttggtagg gccgtgaaat tttgctaaac tgcccatcca 240  
      gaaaatctaag taattttacc aacacgtata gagggccccca cagcttgcgtt atttttgttt 300  
      gtcgaggctc gtctcattca ttattttaa aaggaatttg caacgtcgtg gaaaatgcattc 360  
      tcgaaccacg tcacaatcaa tga 383

50      <210> 4  
      <211> 866  
      <212> DNA  
      <213> Nicotiana tabacum

55      <400> 4

	gaattcgata	gactcactta	aatattagaa	gtgaattacc	tagagttaga	tccaaaacaa	60
	ttatcttgc	cctatccat	caacccttat	ctttcccat	tgattactac	cttgcttacc	120
	tttgttacg	tttcatttag	acaataactt	tagattctta	gttaattgc	gttagaaatt	180
5	atattaaatt	tcaattgttg	gatcatctt	aataccaatc	aagctagaaa	atacaagaat	240
	actgtttaaa	tcaaattccat	gtggatacga	tattatacta	tattatattt	gacttgtgag	300
	cattatttt	gtgtgtttt	tgctcgtaa	agtttggcgt	cgttgcgcag	gattggcaat	360
	caatagtgtt	tgaardatgtt	tttggtgcta	atttaggaat	tagttttat	ttattttattt	420
	tttcttttct	tttcttttcc	ctttctatt	ttatttcctt	tattiagttaa	cttctttca	480
10	agattttttt	tgttagtacct	aacaaggtag	agaagatact	gtagatttt	aactctaaat	540
	gttgtgaaga	tggagtacaa	ccagcctaag	aaaatattt	aatagttgc	agctgaacat	600
	tatcgccggt	cggttatgc	gtttaaatgc	ggtggaaagca	tctaccacccg	cagccctaaag	660
	aaaatatttt	gaatagttag	cagcttgaac	attatccgc	gtcggttatg	tgttttaaat	720
	gccccggaaaa	tcatctacgg	gctaactgtc	aaggcaggat	gtattctcc	tatggttcgt	780
15	attttgagga	gtctcactct	gtttctagtt	cgtacatgt	tgaggattca	tatgggcaca	840
	actctgactc	tggttgggat	gaattc				866

<210> 5  
<211> 998  
<212> DNA

20 <213> Nicotiana tabacum

<400>	5	
25	gaattgtatt attgttaggt gggagagatt ttgtactata tgggtaaaa tcagcgacaa 60 agggcaaat ataccttattt acttttaaaa atagtctaat aatacccttc gttatattat 120 taggttatct ataccttgc agtcatattt tgggttcaaata tatacccctc attaaacgg 180 aggcacacgt gtcatcggtcc tggtggtcaa ttctaaatat ctcctaatta attaaaaaga 240 ctcattaccc atatccgaaa aatattttt aaagcaatat tttttataaa aaaaatggaaa 300 aactgaaatt attttacta aaaattgaaa aaaacgaaaa tagtttttt tcagttttt 360 caaaaaaaaaact attttagaaa aaattgaaa atatttctaa aacaatgtt tttgtaaaaa 420	
30	ctgaaaaaaaaa agaagctgaa aatcaatttt ctaaagaat tttatttgta aaatctggaa 480 aaaactacta aaaactgaaa aaatgaaaat atttttttt ctaattttta caaaaaaaaaac 540 tgctttaaaaa aaagctgaaa atatttctaa aacaatattttt tttgtaaaaa ctaaaaaaaaaa 600 aatattttct tcttttttc agtttttagt taaaaatattttt taagttttt ccagttttt 660 attacttttag aaaattactt ttctgctttt ttttcagttt ttacaaaaat attatttttag 720	
35	aaaatattttt tcagttcttt aaagcagttt tttttgtaa aacttgtaaa aacaatattttt 780 tcgtttttt cagtttttag taaaattttgt ttttagttt tttcagttttt taccaaaaaat 840 aaaatttgct tagaaaattttttt ttttcgggt atgggtatgg tggctttta attaatttagg 900 agatatttttgc aattgtatcaa taggacgtatc acacatgtcc ctccgtttaa atgagggtgt 960 tatttgaatc caaagtatggatc ctgcagccccq qqqqatcc 998	

40  
<210> 6  
<211> 63  
<212> DN

<213> Nicotiana tabacum

45

<400> 6

gaattcga

	atattgaagg taaaatcata tactattttt ctaaaatctc tttaaatgt tccccaatta 120
	tctgatttct atattgctct taaaatgtcac tcaaccttag atcaacaac atataactt 180
50	cccgatcat aagagattgc ggcattatgg cttaaaaaaa atcctagaga cacatctgc 240
	ccacatattt gaattttatac ccacagtaat agagctcggt tagtacatta ttattatgg 300
	tgttacgatc cgttgcgtt tccattttaa ttcccttcgt gtggaaatgg atgacattgt 360
	ggaattaaaa aaatttattca gacaaaaaat tcgacgaaac gtagagctt ctgcgaacat 420
	gaacaattgc ccagtatatac aaatacgtgt tcagttgtat gattcccttgat tatggaaagat 480
55	gaatcactac aaagaggaaaa acgaaaaaga gatacagtgt ctgtcgaga gtattattgt 540
	tacaatttc aactaagaaaa taatgaaaca aatgaagtgt tacattgtgg gagaatattc 600



tatataattca cataacttata gtacaaaactt tgtttctatt tataaaagaga aaaagaaaatt 240  
 ttacaaaaaa caaatatatt tgcttcattt taatttagtag ttttattaag caagctata 300  
 aagctc 306

5 <210> 10  
 <211> 685  
 <212> DNA  
 <213> Nicotiana tabacum

10 <400> 10  
 gaattccgtg gttttgcac ggctcgctcaa ttgtcatatt tggctcattt atctgat 60  
 taaacaatta agaacttata tgcaaattta acttttaaaa ccgttttat cattatttat 120  
 ttatcacaa attacaacgt cgtaaaaagg catctcgAAC cacGCCcaaa ccAGTGCACA 180  
 cgtgatttgt tgacgcattt tggacttcgt caagatcgtg atttgggtt cataatgt 240  

15 caccggat ttaagaaaat aactttata aatattgcgc caaaatacta cgcgttatga 300  
 tactatttagt gtaggcatttgaatattttact aaatcgccc tctcggaatc taggtat 360  
 ctatattaa aaaaaataag atggggcct gcaattttt attatattat atttatttat 420  
 tttagcga agatccctcc ctatattttaat gataaccctt taatgactac atctttat 480  
 ttactaagtt tgtctataat tatgaagtca atctctacat acataaaaat aacatattaa 540  

20 ttactaattt aaaaacaata ttaatggaaa gtaatattac taaaattata attacaaca 600  
 acatgaaatt gtcacaaaat aaaaataaa aactaattat cccatagttt gattaaaatt 660  
 catattgtt gtagactt agttt 685

<210> 11  
 25 <211> 899  
 <212> DNA  
 <213> Nicotiana tabacum

<400> 11  
 30 aagctttaaa aggaagagag ccacaattttt ctttgaccctt ctttctctcc tagccactaa 60  
 gatatacgt actggtcataa aagagcatat ttatagctca aaattttgcc ttttctgtt 120  
 gtaaaacgtga ttgtttctta ctggattct tggcttatattt attacggga gaaaagagca 180  
 atttgcattc tcctaaatct ttatatttctt ggtgaaaat tggctttaa ttggctggga 240  
 atttattttt agatgctaca accttgacaa acacctaaga atattttatg gacaatggct 300  

35 tggctttga gtactggttt ttctgtttctt ggtccctgtt tcaacggcc acggcaagag 360  
 tctcgtcgatc attggcccttc gattggcact ctgcaactt aagattttgc atccagagaa 420  
 ttcttaggca aacccttgggat ttagcatca gatcatattt gtagacttacca ttggatgtt 480  
 gaacgactttt ctgtatgtaa atctctgata catttgcctt tggctttata ctgggtgtt 540  
 tcatgttttcaatttgcattt taaatttttcaat gatcatcaat catttataag tatttattct 600  

40 aatgatttttta ggcacaagta tcaatcgctg ctcagagatg gtgggagaag acccttaaac 660  
 ccaacatggt agagatcaat tcgcaacac aacttgcattt ttcatttataa aagctgtt 720  
 atagattggat ctaatttgcattt ttcatttcttc ttagctgttag aggtttcaag actttacatc 780  
 ctaaggtttagt atatatagca atccccctaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaccaacaa 840  
 ctacatcgatc atccataagca agtttagggtt aactatatgc atcatcaacta gacggatcc 899

45 <210> 12  
 <211> 999  
 <212> DNA  
 <213> Nicotiana tabacum

50 <400> 12  
 aagcttaact ttactcataat tgctttctttt agggaaagcgt ctgtttttttt gaccatccctc 60  
 taaatttctc atgaatcttc ttctgttgc cactctgtt aatgtcatca tggctgaaac gaaatctgaa 120  
 attgtcatca tggctgactat tatccatca ctcgtctt aatgtcatca tggctgaaat 180  

55 tggcttgcaccag cccataactga ttttttttttt tttgggtctt aactttccct tccggtagtc 240  
 ggttggagtc atgaacttat ttctgttgcattt gaggatatgc ctgtatggcc tatactctt 300

tggtgtctca aggccgtcga cctctcatct ttccttcaa ttgactatacg actctgtaat 360  
 actgtcatct ttggatcta ccgttgcct ccatgtatca tatcttactc ataatgcctc 420  
 attaactatt ttcttatttc cgcataacat ttatgtctat cactttattc tgaaaactcg 480  
 aacaagacat tctttcgtt ttagatcccc ttgtctccat ccagtggttc ttcggggac 540  
 5 ttaacgttct cgcttccta gggaggcgag ccacactaag gtaatattta tcccttctag 600  
 gcttccgtg cctatcttct gagatatttt ttcatgcta atattcacat ctaattgtaa 660  
 ttttctagag tgccatct gggtgcctca caagaagagc tattagcatc tttgtat 720  
 ccttcgaaaa tgtcaactaa cacaacacaa tccattcacc attttggtt actctaacct 780  
 cagtcggata ctaatatctt gtcattttat taaactacac atgttagccc ccaatagat 840  
 10 ataaactaaga tgggtgtgcc caattctaca tacatctgtt actgttgaaa gtaagtgcga 900  
 atgcttttat ttttcgtcc gagttaaaaa taccgataat ctatattaac tgggtacctc 960  
 gtacccttct catcttctc ctttacttg tgaagctt 999

<210> 13  
 15 <211> 1499  
 <212> DNA  
 <213> Nicotiana tabacum

<400> 13  
 20 aagcttggaaa aagaagaatt aaggcttgct ttcttaattt taaaaaaaata aaaatttattt 60  
 tgaactatct atactatattt aaaggcacga aaaccctatc gaaatgtcg tcgcctttt 120  
 tacccttaa aaataatttt acattagaca aaatagtcat tttactattt ttccataat 180  
 ataggattttt aaaattaattt taactttggc tattaaacat ttcttataaa cttgaaat 240  
 gtaaaactcc taatatttag aaattaattt aacataacca aggattttt tattcgtaat 300  
 25 aactctaata tggtatccaa atcgtcttag aactcttta cctctaataa gtaaaagtac 360  
 ttcttaataaa ttcatataact tttctctct tctccgatct ctctttgctc ttctttttat 420  
 gtatcccttc ctttctaata gcctttatg agaagtaaac tttaggggtt ggccccccct 480  
 ccccccacaa ttatatagtt tctactctag ttgttggaaat ataattcaaa ttcttaataa 540  
 attgacggtg acattggattt ttactttgtt gaagagaattt agattctcg ttttagt 600  
 30 tcggtttagta attgtatgtt cattttttt actctataat agagatgcaa ttttattttt 660  
 gcattttggg atcaaattgtt aatgcagtca tatattgtt tcataatgtt ttggatatt 720  
 gttggttattt taacttagaaa tagacttctt atttcatattt tattgtttaaa atcccttattt 780  
 ggagatgaat tatttgttca ccgatttagaa gttgatagtc gctttgttt tagaagaaat 840  
 tttaccgttag accaagttaa ggagtttttag aagcactttt catgggagca ttgtgtat 900  
 35 ttatqgctt atcaaataata gtttttgaag attcagagag ccaagaaaag ctagaaccca 960  
 agaacttagga agttagatgtt attcacaata ccataacgtg atataaaaact ttttattgtt 1020  
 actcaaatcg gtaatattttt ttgctttatg cttaatcgat aaatttatttt ttatattgtt 1080  
 ttagttatag gaggctcaca aagttggaaat taattaaaat atcatattttt gtattgtt 1140  
 aattttagaa atagaattt gtaaaaaatc actttaaattt ttatccat atccagaagg 1200  
 40 attatgggtt ctggcatatgt ttgttggaaat atttgaatca gggtaaaatg atgttgtaat 1260  
 ttttattttt ttataggcat ttgttgcctt tgattttttt gttgtcattt tatttttattt 1320  
 ttgttggaaatg tataatatgt ttgttggaaat atatagataa tcaattttat aagaaattt 1380  
 caacaattac acaaggataa agtctacaat atgcgagtaa aatttgattt aacctaggat 1440  
 gtcataattt atgcataattt tatttcaatg tggttattat acatctattt tattatattt 1499

45 <210> 14  
 <211> 10  
 <212> DNA  
 <213> Nicotiana tabacum

50 <400> 14  
 aataaayaaa

55 <210> 15  
 <211> 10  
 <212> DNA

<213> Nicotiana tabacum  
 <400> 15  
 ttwtwtwtt 10  
 5 <210> 16  
 <211> 11  
 <212> DNA  
 <213> Nicotiana tabacum  
 10 <400> 16  
 wtttatrtt w 11  
 <210> 17  
 15 <211> 15  
 <212> DNA  
 <213> Nicotiana tabacum  
 <400> 17  
 20 gtnwayattn atnnrr 15  
 <210> 18  
 <211> 18  
 <212> DNA  
 25 <213> Nicotiana tabacum  
 <400> 18  
 tgtaaaacga cggccagt 18  
 30 <210> 19  
 <211> 18  
 <212> DNA  
 <213> Nicotiana tabacum  
 35 <400> 19  
 caggaaaccg atatgacc 18  
 <210> 20  
 <211> 1103  
 40 <212> DNA  
 <213> Nicotiana tabacum  
 <400> 20  
 tcgattaaaa atcccaatta tatttcaga taaaatcaa ccataactca ttttgtttaa 60  
 45 gcttggtttg gttttatatt tatataagtt ttatataata tgcccttaag actttttata 120  
 gaattttctt taaaaaatat ctggaaatat ttgcgactct tctggcatgt aatatttcgt 180  
 taaatatgaa gtgcctccatt ttatataact taaaataatt ggttgcacgt tcactttctt 240  
 atcaagtgtt actaaaatgc gtcacatctc ttgttcttcc atattcatat gtcaaaatct 300  
 atcaaaaattc ttatataatct ttgttcaatt tgaagtggaa ttgcataat taaaattaa 360  
 50 atagaacata tcattattta ggtatcatat tgatttttat acttaattac taaaatttggt 420  
 taactttgaa agtgtacatc aacgaaaaat tagtcacacg actaaaataa ataaatataca 480  
 tgtgttatta agaaaaattct cctataagaa tattttataa gatcatatgt ttgtaaaaaaa 540  
 aattaatttt tactaacaca tatatttact tatcaaaaaat ttggcaaaac cgaaccaatc 600  
 caaccgatat agttggtttg gtttgatttt gatataaacc gaaccaactc ggtccatttg 660  
 55 cacccttaat cataatagct ttaatatttc aagatattat taagttacg ttgtcaatat 720  
 cctggaaatt ttgcaaaaatg aatcaagcct atatggctgt aatatgaatt taaaaggcagc 780

tcgatgtggt ggtaatatgt aatttacttg attctaaaaa aatatcccaa gtattaataa 840  
tttctgtcgag aagaagggt agctacgatt tacagcaaag ccagaataca aagaaccata 900  
aagtgattga agctcgaat atacgaagga acaaataattt ttaaaaaaat acgcaatgac 960  
ttggaacaaa agaaagtgtat atatttttg ttcttaaacaca agcatcccct ctaaagaatg 1020  
5 gcagttttcc ttgcgtatgt actattatgc tcccttcgtt acaaaaattt tggactacta 1080  
ttgggaaccttctgtaaaa tag 1103

<210> 21  
<211> 838  
10 <212> DNA  
<213> Nicotiana tabacum

<400> 21  
aagcttacat ttatgttag ctggtgact gacgccagaa aatgtggtg atgcgttag 60  
15 attaatggcg ttatgggtgt tgatgtaaac ggaggtgtgg agacaatgg tgtaaaagac 120  
tctaacaacaaa tagcaaattt cgtaaaaaat gctaagaaat aggttattac tgagtat 180  
ttatTTAAGT attgtttgtg cacttgctg caggccttt gaaaagcaag cataaaagat 240  
ctaaacataa aatctgtaaa ataacaagat gtaaagataa tgctaaatca ttggcttt 300  
tgattgattt tacagaaaaa tatacatcgc agggggttga cttttaccat ttcaccgcaa 360  
20 tggaaatcaa ctgttgaag agaatgttca caggcgcata cgctacaatg acccgattt 420  
tgctaggctt ttctcggtct tgcaaacaac cgccggcagc ttagtatata aatacacatg 480  
tacataccctc tctccgtatc ctgcgtatc ttttcttgc tttatcgctt ttgcgtgt 540  
aaaactttat cacacttac tcaaaatacac ttattaaccg cttttactat tatcttctac 600  
gctgacagta atataaaaca gtgacacata ttaaacacag tggtttctt gcataaaacac 660  
25 catcagccctc aagtgcgtcaa gtaaagattt cgtgttcatg cagatagata acaatctata 720  
tggtgataat tagcgttgcc tcatcaatgc gagatccgtt taaccggacc ctgtgcact 780  
taccccacgt tcggtccact gtgtgccgaa catgctcctt cactatTTA acatgtgg 838

<210> 22  
30 <211> 9  
<212> DNA  
<213> Nicotiana tabacum

<400> 22  
35 aatataattt

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/16344

A. CLASSIFICATION OF SUBJECT MATTER			
IPC 6	C12N15/82	C12N5/10	A01H5/00
			C12O1/68

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12N A01H C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	WO 97 27207 A (NORTH CAROLINA STATE UNIVERSITY) 31 July 1997 see page 2, line 9 - line 34 see page 6, line 16 - page 7, line 29 see page 12, line 27 - page 13, line 28 see page 21, line 3 - page 25, line 21 --- -/-	20,21 1-19

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

15 December 1998

Date of mailing of the international search report

22/12/1998

Name and mailing address of the ISA  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl  
Fax: (+31-70) 340-3016

Authorized officer

Montero Lopez, B

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 98/16344

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GEORGE C. ALLEN ET AL.: "High-level transgene expression in plant cells: Effects of a strong Scaffold Attachment Region from Tobacco" THE PLANT CELL, vol. 8, May 1996, pages 899-913, XP002072398	20, 21
A	see page 900, left-hand column, paragraph 2 see page 908, right-hand column, paragraph 1 - page 909, left-hand column, paragraph 2 ---	1-19
A	WO 94 07902 A (NORTH CAROLINA STATE UNIVERSITY) 14 April 1994 see page 2, line 5 - page 3, line 21 see page 4, line 1 - page 5, line 17 see page 10, line 5 - page 11, line 5 -----	1-19

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International Application No

PCT/US 98/16344

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9727207	A 31-07-1997	US 5773695 A		30-06-1998
		AU 2246997 A		20-08-1997
WO 9407902	A 14-04-1994	AU 673859 B		28-11-1996
		AU 5165593 A		26-04-1994
		CA 2147006 A		14-04-1994
		EP 0663921 A		26-07-1995
		US 5773689 A		30-06-1998